CONTRACT NAS5-32605
EOSDIS IV&V

EOSDIS CORE SYSTEM (ECS) MODELING ASSESSMENT REPORT

(Deliverable 0506)

July 31, 1995

Prepared By:

INTERMETRICS
6301 Ivy Lane Suite #200
Greenbelt, MD 20770

Prepared for:

NASA Goddard Space Flight Center
Code 505
Greenbelt, MD 20770

CONTRACT NAS5-32605
EOSDIS IV&V

EOSDIS CORE SYSTEM (ECS) MODELING ASSESSMENT REPORT

(Deliverable 0506)

July 31, 1995

REVIEWED BY:	PREPARED BY:
Debbie Izumi Task Leader	Larry McGoldrick Task Member
RECIEVED BY:	PREPARED BY:
Lee LaCoste Document Log Manager	Layne Miller Task Member
APPROVED BY:	PREPARED BY:
Ron Cariola Program Manager	Lionel Mitchell Task Member

INTERMETRICS
6301 Ivy Lane Suite #200
Greenbelt, MD 20770

TABLE OF CONTENTS

Section
Page

1. EXECUTIVE SUMMARY 1-1
2. INTRODUCTION Error! Bookmark not defined.
2.1 Purpose Of The Report 2-1
2.2 Objective Of The Analysis 2-1
2.3 Scope Of The Analysis 2-1
2.4 Background Information 2-1
3. USER MODEL 3-1
3.1 Analysis Tasks Performed 3-1
3.1.1 Background 3-1
3.1.2 *Who* Are the Users (Demographics) 3-1
3.1.3 *What* Are They Interested In (Products) 3-1
3.1.4 *How* Will They Use the System (Services Invoked) 3-2
3.2 Constraints Affecting The Analysis 3-2
3.3 Results 3-2
3.3.1 Discussion of Results 3-2
3.3.2 Identified Problems 3-7
3.3.3 Potential Issues 3-7
3.4 Conclusions 3-8
3.4.1 Technical Integrity 3-8
3.4.2 User Satisfaction 3-8
3.4.3 Trends and Projections 3-8
3.5 Recommendations 3-9
3.5.1 Areas Requiring Further Analysis 3-9
3.5.2 Solutions To Important Problems 3-9
3.5.3 Risk Management 3-10
4. PRODUCTION MODEL 4-1
4.1 Analysis Tasks Performed 4-1
4.1.1 Background 4-1
4.1.2 TB Structure and Methodology 4-1
4.1.3 TB Self-Consistency 4-1
4.2 Constraints Affecting The Analysis 4-2
4.3 Results 4-2
4.3.1 Discussion of Results 4-3
4.3.2 Identified Problems 4-8
4.3.3 Potential Issues 4-8
4.4 Conclusions 4-8
4.4.1 Technical Integrity 4-8
4.4.2 User Satisfaction 4-9
4.4.3 Trends and Projections 4-9
4.5 Recommendations 4-9
4.5.1 Areas Requiring Further Analysis 4-9

4.5.2 Solutions To Important Problems 4-10
4.5.3 Risk Management 4-10
5. PERFORMANCE MODEL 5-1
5.1 Analysis Tasks Performed 5-1
5.1.1 Evaluation Areas 5-1
5.1.2 Evaluation Metrics 5-2
5.2 Constraints Affecting The Analysis 5-4
5.3 Results 5-4
5.3.1 Discussion of Results 5-4
5.3.2 Identified Problems 5-15
5.3.3 Potential Issues 5-17
5.4 Conclusions 5-17
5.4.1 Technical Integrity 5-17
5.4.2 User Satisfaction 5-18
5.4.3 Trends and Projections 5-19
5.5 Recommendations 5-19
5.5.1 Areas Requiring Further Analysis 5-19
5.5.2 Solutions To Important Problems 5-19
5.5.3 Risk Management 5-21
6. COST MODEL 6-1
6.1 Analysis Tasks Performed 6-1
6.1.1 COTS HW and SW Estimation 6-1
6.1.2 Custom SW Estimation 6-2
6.1.3 Operations and Maintenance Cost Estimation 6-2
6.1.4 Model Interfaces and Design 6-2
6.2 Constraints Affecting the Analysis 6-2
6.2.1 COTS HW and SW Estimation 6-3
6.2.2 Custom SW Estimation 6-3
6.2.3 Operations and Maintenance Cost Estimation 6-3
6.3 Results 6-3
6.3.1 Discussion of Results 6-4
6.3.2 Identified Problems 6-7
6.3.3 Potential Issues 6-8
6.4 Conclusions 6-8
6.4.1 Technical Integrity 6-9
6.4.2 User Satisfaction 6-9
6.4.3 Trends and Projections 6-10
6.5 Recommendations 6-10
6.5.1 Areas Requiring Further Analysis 6-10
6.5.2 Solutions to Important Problems 6-11
6.5.3 Risk Management 6-11
APPENDIX A: USER MODEL ANALYSIS DETAIL A-1
A.1 User Demography Tables A-1
A.2 EGSUS Products A-2
APPENDIX B: PRODUCTION MODEL ANALYSIS DETAIL B-1

APPENDIX C: Performance Model Analysis Detail C-1
C.1 System Representation Evaluation C-2
C.2 Workload Representation Evaluation C-3
C.3 Performance Statistics Evaluation C-6
C.4 Model Structure Evaluation C-7
APPENDIX D: COST MODEL ANALYSIS DETAIL D-1
D.1 Analysis Methods D-1
D.1.1 COTS HW and SW Estimation D-1
D.1.2 Custom SW Estimation D-2
D.1.3 Operational Cost Estimation D-4
D.2 Analysis Results D-5
APPENDIX E: LIST OF REFERENCES E-1
APPENDIX F: TOOLS AND DATA BASES UTILIZED F-1

This page intentionally left blank.

TABLE OF EXHIBITS

Exhibit
Page

EXHIBIT 1-1: Model Maturity Levels 1-4
EXHIBIT 2-1: The ECS Modeling Context **Error! Bookmark not defined.**
EXHIBIT 2-2: The EOSDIS Users (Researchers)—Two Perspectives 2-3
EXHIBIT 2-3: The EOSDIS User Services in the User Scenarios 2-4
EXHIBIT 2-4: The CDR Mission Baseline 2-4
EXHIBIT 3-1: User Model Maturity Table 3-2
EXHIBIT 3-2: Discipline choices from the 288 reliable records 3-3
EXHIBIT 3-3: Discipline choices from previous work 3-4
EXHIBIT 3-4: Discipline choices from the EGSUS Database 3-4
EXHIBIT 3-5: USGCRP Research Area Preferences Common to the EGSUS Database and the NASA HQ. Survey 3-5
EXHIBIT 3-6: USGCRP Research Area Preferences from [3], for comparison 3-5
EXHIBIT 3-7: User Scenario Matrix, Showing Respondents Discipline Area 3-6
EXHIBIT 4-1: Contents of the CDR Technical Baseline 4-2
EXHIBIT 4-2: Production Model Maturity Table 4-3
EXHIBIT 4-3: From the AHWGP Process Description Details v2.1 4-4
EXHIBIT 4-4: From the AHWGP-Based TB File Description Details v2.1 4-4
EXHIBIT 4-5: From the AHWGP-Based TB Processing Timelines v2.1 4-5
EXHIBIT 4-6: From the AHWGP-Based TB Volume Timelines v2.1 4-5
EXHIBIT 4-7: CDR Data Volume Summary Spreadsheet 4-6
EXHIBIT 4-8: Aster Processing Load; Bottom Row Shows Totals 4-7
EXHIBIT 4-9: ADEOS II (SWS) and RADAR ALT (DFA) Summary 4-7
EXHIBIT 4-10: CERES Choices from the Volume Timelines Spreadsheet 4-8
EXHIBIT 5-1: Performance Model Maturity Table 5-6
EXHIBIT 5-2: Top Level BONeS Model Diagram 5-7
EXHIBIT 5-3: Computer Resources Types 5-8
EXHIBIT 5-4: Resource Metrics 5-11
EXHIBIT 5-5: Model Results Completed 5-12
EXHIBIT 5-6: Model Studies Planned 5-13
EXHIBIT 5-7: Validation Results 5-14
EXHIBIT 5-8: Problem Areas and Characterizations 5-17
EXHIBIT 6-1: Cost Model Maturity Table 6-4
EXHIBIT A-1: User Discipline Analysis: EGSUS Only (See exhibit 3-2) A-1
EXHIBIT A-2: User Discipline Analysis: EGSUS and NASA HQ Survey (See exhibit 3-4) A-1
EXHIBIT A-3: User USGCRP Research Area Analysis (See exhibit 3-5) A-1
EXHIBIT A-4: EGSUS Product Choices A-3
EXHIBIT C-1: Evaluation Decomposition C-1
EXHIBIT C-2: Subsystem Evaluation C-2
EXHIBIT C-3: Resource Evaluation C-2
EXHIBIT C-4: Subsystem Interface Evaluation C-2
EXHIBIT C-5: Push Workload Evaluation C-3
EXHIBIT C-6: Pull Workload Evaluation C-4
EXHIBIT C-7: Reprocessing Workload Evaluation C-5
EXHIBIT C-8: V0 Workload Evaluation C-6
EXHIBIT C-9: System Overheads C-6
EXHIBIT C-10: Resource Metrics Evaluation C-6
EXHIBIT C-11: Performance Requirements Evaluation C-6

EXHIBIT C-12: Performance Results Evaluation C-7
EXHIBIT C-13: Module Modules Evaluation C-7
EXHIBIT C-14: Model Data Structures Evaluation C-7
EXHIBIT C-15: Model Parameters Evaluation C-7
EXHIBIT D-1: Parameters Used in Estimating the Average Cost per Man Year D-5
EXHIBIT D-2: Cost Modeling Requirements Satisfaction Matrix D-6
EXHIBIT F-1: Tools and Databases Used F-1

This page intentionally left blank.

1. EXECUTIVE SUMMARY

This technical analysis report (TAR) presents the results of an independent assessment, conducted by the EOSDIS IV&V team during the period from 13 February 1995 to 31 July 1995 to evaluate the modeling activity being performed by the ECS development contractor Hughes Applied Information Systems (HAIS). It is an update to our previous Modeling Assessment TAR [3]. The objective of this independent analysis is to provide the ESDIS Project with objective insights into the validity of the underlying assumptions and predictive quality of the models used by HAIS to support the ECS Critical Design Review (CDR) evaluation by the Government. The results presented in this report focus on the state of the models (i.e., the level of maturity attained) at the time they were used to generate predictions in support of the CDR. Detailed analyses of the system-level modeling activities associated with user loading and demand (User Model), production loading (Production Model), system architecture and performance (Performance Model), and cost estimation (Cost Model) have been conducted. A summary of findings for each of these areas follows.

User Model

The User Model, also referred to as the user pull model, describes the interaction of Earth-science researchers and other users with the EOSDIS. Our analysis focuses on the characterization of user demographics (who are the users) and products and services (what do they want from the system). Information was taken from two primary sources, both developed by HAIS: user scenarios, and the preliminary EOSDIS General Science User Survey (EGSUS) database. The characterization of user demography and product demand is still in progress by the ECS contractor. Therefore, the results presented here are preliminary and based on a snapshot in time of ongoing user modeling activities. The analysis was further constrained by delayed availability of the preliminary EGSUS results and the CDR Technical Baseline.

To date, the HAIS characterization of user demography appears reasonable; however, the maturity level of this activity is somewhat limited. The compilation and analysis of the EGSUS results are still in progress. It is questionable whether the user scenarios adequately represent the expected distribution of science users; this could potentially result in incorrect estimates of user service and data requirements that are key inputs into the Performance Model. Continued analysis of the EGSUS and the Ad Hoc Working Group on Consumers (AHWGC) survey should provide the additional necessary information.

The frequencies of occurrence of service invocations and products requests are needed inputs to the Performance Model. This information, derived form the user scenarios is detailed in some areas (many service types) and sparse in other areas (only 27 scenarios). Similar information derived from the EGSUS database will contain fewer service types, but the product list will be broader. Moreover, there are 288 mini-scenarios in the preliminary database we analyzed. This is an area in which the AHWGC survey will play a crucial role, since the level of detail of the products and services requests is expected to be greater. This part of the user modeling activities has reached a level of limited maturity due to the limited analysis that has been performed to date by HAIS.

Production Model

The Production Model (i.e., data push model) describes the science-product generation process (nominally, Level 0 input transformation to Level 1, 2, 3, and 4 products). It is intended to predict the steady-state and exceptional processing necessary to deliver trusted science data to the EOSDIS archives when required. Results from the Production Model formulate the basis for determining the inputs to the Performance Model, and ultimately, information for the ECS system developers. The CDR Technical Baseline (TB), prepared by the ECS Project, is the primary information source for push load analysis; it includes information on EOSDIS products and their file sizes, processing requirements, and frequencies and timelines for production. Our analysis focuses on the structure and consistency of the TB and the methodology used to produce inputs for the Production Model. Due to delayed availability of the CDR TB (available to IV&V on June 27), our analysis is limited to a preliminary assessment of estimated production demands for some specific instruments in the AM-1 and TRMM missions.

Based on our preliminary analyses, the structure of the CDR TB appears to contain the necessary information for the HAIS performance modeling staff to derive the input for the Performance Model. The process by which the TB information can be converted into the required Performance Model input tables is sound and reasonably complete. However, some of the instruments do not have information specified to the appropriated level of detail to adequately model their production processes. Inconsistencies were found among the various spreadsheets contained in this TB, which could lead to spurious Performance Model outputs. It is essential for the TB to be complete, consistent, and error-free to ensure optimal inputs to the Performance Model, which in turn drives ECS design and development. For the reasons stated above, the production modeling activities have achieved a somewhat limited maturity for ASTER and limited maturity for CERES.

Performance Model

The Performance Model is intended to provide a basis for evaluating the performance of proposed ECS configurations capable of supporting user and production demands (as predicted by the User and Production Models), and evaluating sensitivities of various system designs to predictive uncertainties and evolving requirements. This model is implemented using the Block Oriented Network Simulator (BONeS) discrete-event simulation modeling tool. The IV&V analysis focused on the evaluation and validation of the following: representation of the system design, representation of the system workloads, methods of statistics collection and performance metrics quantification, and the overall model structure.

Our analysis was constrained by limited available simulation results. The results available focus primarily on ingest and production of standard products for TRMM and AM-1 instrument loads. Lack of CDR-level documentation for the Release A SCDO detailed design was a secondary constraint. Results reported here are based on the state of the model and supporting information as of 30 June 1995.

Overall, the Performance Model is at a somewhat limited level of maturity. Although significant improvements have been achieved since PDR, the model lacks functionality that should be available for CDR. The system representation was evaluated for the inclusion of the appropriate subsystems, the distributed sites, and computer and communication resources of the design. Our analyses show that some of the hardware and software components of some of the subsystems are not represented (e.g., Data Management); others are minimally represented (e.g., Planning). The BONeS model modules and parameters were examined for complete and accurate representation of the workload frequencies and flows through the system resources. Workload representation in the model is lacking in several areas (e.g., reprocessing, non-EOS push data, DAO/DAS). While the model has a high level of maturity for resource performance metrics, performance statistics for quantifying response times are not now collected. The statistical analysis methodology used to analyze the model data is sound; however, more results should be generated and validated to support the CDR. In general, the ECS Performance Model is well structured. It is modularized and parameterized to a high degree so that modifications can be made with minimal impact to the existing model components.

Cost Model

The Cost Model provides resource estimates required to develop and operate ECS architectural alternatives (as partially derived from the Performance Model) within schedule constraints. The IV&V analysis focused on a reevaluation of the Cost Model, implemented as a collection of three independent estimation models: Commercial Off-The-Shelf (COTS) hardware and software models, used to estimate hardware, software, and procurement costs, and perform cost impact analysis; a custom software model, used to estimate size, level of effort, schedules, and associated costs for custom developed software; and an operations and maintenance model, used to estimate personnel costs associated with ECS operations and maintenance.

Information to support the Cost Model analysis was obtained exclusively through interviews and discussions with HAIS. With the exception of the HAIS Interactive Cost Model (ICM), our analysis was severely constrained by not having the models and pertinent cost information in hand. Overall, we consider the Cost Model to be somewhat limited maturity due to the lack of a lifecycle cost model, trade studies that are based on restricted input, use of conservative parameters, and a custom software estimation approach that is untested. Given the current implementation of the Cost Model, there is no integrated mechanism to estimate lifecycle costs and subsequently perform what-if analyses. Without this capability, it is difficult to analyze the lifecycle benefits or drawbacks of implementing alternative solutions. The lack of an integrated lifecycle cost model could result in trade analyses based on only a subset of cost data, such as COTS hardware and software costs. The results can be misleading and actually lead to design decisions that increase, rather than decrease, costs. Of equal concern is the potential that some of the modeling parameters may be too conservative. Although some degree of conservatism may be warranted in the early stages of the lifecycle, overly conservative estimates can have negative implications such as the unnecessary de-scoping of systems. The custom software estimation method used is unproven in the industry and potentially underparameterized. Performing an independent second estimate using a different model is strongly recommended.

Conclusion

We have noticed significant progress in some of the user, production, and performance modeling areas. However, due to repeated unsuccessful attempts to obtain detailed cost information from the ECS contractor, we have been unable to determine changes and improvements in the cost modeling area. Although we have noticed increased maturity in most of the modeling areas, overall, the models are all of somewhat limited maturity with respect to what should be completed to meet CDR requirements. Deficiencies identified could potentially lead to inaccurate or incomplete predictions of performance and costs. Based on our assessment of the accomplishments to date, the modeling efforts could yield a viably mature set of user, production, and performance models after the CDR time period.

Model Class	Model Element	CDR Maturity	Maturity Metrics
User Model	Overall	2	
	User Profile (Who)	2	0: Not Addressed
	User Needs (What)	1t	1: Limited Maturity
	User Access (When)	1t	2: Somewhat Limited Maturity
	User Services (How)	1t	3: Fully Mature
	Output to Performance Model		
Production Model	Overall	2	Blank: not analyzed by IV&V
	AM-1	2	
	TRMM	1	
	Other Missions		
Performance Model	Overall	2	?: Estimated based on limited IV&V inputs
	System Design Representation	2	
	Workload Parameters	2	
	Performance Statistics	2	
	Model Structure	3	
Cost Model	Overall	2?	
	COTS H/W and S/W	2?	
	Custom Software	2?	
	Operations & Maintenance	2?	
	Model Interfaces (inc. Perf Mod)	1?	

EXHIBIT 1-1: Model Maturity Levels

To The Reader: If your planned reading of this document is limited to the Executive Summary, please consider also reading Section 2.4 (Background Information) to *gain a full understanding of the context within which we arrived at our conclusions and recommendations.*

This page intentionally left blank.

2. INTRODUCTION

This section provides introductory information pertinent to this IV&V *EOSDIS Core System (ECS) Modeling Report*. The purpose of the report, objectives and scope of the analysis, and relevant background information follow.

2.1 Purpose Of The Report

The purpose of this technical analysis report (TAR) is to document to results of an independent assessment of the ECS contractor's activity that was conducted by the EOSDIS IV&V team over the period 13 February 1995 to 31 July 1995. This TAR documents identified problems and potential issues, including their relative severity and possible adverse implications for employing the models to reliably predict ECS performance and cost (development and operation) estimates. This TAR is the third in a series of reports and follows the preliminary report [3] submitted in February, 1995.

2.2 Objective Of The Analysis

The objective of this analysis is to independently assess the ECS modeling activity being performed by the ECS development contractor (Hughes Applied Information Systems [HAIS]). The intent is to provide the ESDIS Project with objective insights into the validity of the underlying assumptions and predictive quality of the models to support the ECS Critical Design Review (CDR).

2.3 Scope Of The Analysis

This analysis examines system level modeling activities associated with user and production loading, system architecture and performance, and cost estimation. The scope is limited to the state of the models (i.e., the level of maturity attained) at the time they were used to generate predictions in support of CDR. The analysis does not examine subsystem level modeling activities (such as DADS) that indirectly affect the quality of system level models by providing parametric drivers, typically in the form of subsystem response characteristics.

2.4 Background Information

The ECS modeling activity focuses on the formulation and implementation of four classes of interrelated models: the User Model, Production Model, Performance Model, and Cost Model. Exhibit 2-1 illustrates the context within which the models interact to generate predictions of performance and cost. The follow discussion is keyed to this exhibit. The discussion does not address the correctness of the context; that is the subject of later sections of this TAR.

EXHIBIT 2-1: The ECS Modeling Context

The User Model describes the interaction of Earth-science researchers and other users with the EOSDIS, and is also referred to as the user pull model. This model is intended to predict what the user needs and loading will be for the EOSDIS and how the demands will vary, over time, as more capabilities come on-line and users gain more experience with the system. The different types of users (see Exhibit 2-2) have been surveyed with somewhat limited utility so far. A new survey, the EOSDIS & General Science User Survey (EGSUS), has been distributed by HAIS via the World Wide Web (WWW) to better identify what set(s) of science data products, from a list of products available over time, are of interest to the users and how and to what extent they may interact with the EOSDIS. The users' stated interaction scripts were compiled into scenarios. The results of this compilation are then translated into user services (see exhibit 2-3) that form the basis for the Performance Model's pull workload characterization. The temporal details are derived from the relative access frequency distributions of services (derived from the user scenarios) within epochs of time that closely map to mission milestones. These spreadsheet-based workload distributions are finally translated, manually, into the Performance Model's input parameter tables. As the ECS matures and current Version 0 (V0) user-interaction measurements become available, it should be possible to calibrate the model with real-world information and generate workload predictions with progressively improving confidence.

The ECS Contractor View: Traditional Disciplines	**The USGCRP View[†] Global Change Research Areas**
Atmosphere (atm)	Climate and Hydrologic Systems (chs)
Cryosphere (cryo)	Biogeochemical Dynamics (biodyn)
Land (land)	Ecological Systems and Dynamics (eco)
Ocean (ocean)	Human Interactions (humint)
	Earth System History (hist)
	Solid Earth Processes (solid)
	Solar Influences (solar)

[†] U.S. Global Change Research Program

EXHIBIT 2-2: The EOSDIS Users (Researchers)—Two Perspectives

User Service Class	Number of Discrete Functions/Services
Search	4
Manipulate	21
Inspect	13
Archive	6
Ingest	1
Produce	1
Other	3
Total	49

EXHIBIT 2-3: The EOSDIS User Services in the User Scenarios

Mission	Instruments	Launch Date
TRMM	CERES, LIS, VIRS, PR, TMI	17 August, 1997
Landsat 7	ETM+	May, 1998
FOO	COLOR	July, 1998
EOS AM-1	ASTER, CERES, MISR, MODIS, MOPITT	30 June, 1998
METEOR	SAGE III	August, 1998
ADEOS II	SeaWinds	February, 1999
ALT RADAR (CNES or GFO)	MR, DFA	March, 1999
ACRIMSAT	ACRIM	June, 1999
Space Station	SAGE III	June, 2000
EOS PM-1	AIRS, AMSU, CERES, MIMR, MODIS, MHS (NOAA)	December, 2000
CHEM	HIRDLS, MLS, CII, TES	December, 2002
ALT LASER	GLAS	July, 2003

EXHIBIT 2-4: The CDR Mission Baseline

The Production Model describes the science product generation processes (nominally, Level 0 input transformation to Levels 1/2/3/4 product—the data push model). This model is intended to predict the steady-state and exceptional processing necessary to deliver trusted science data to the EOSDIS archives when required by the users. The Mission Baseline (spacecraft and instrument manifest) from the CDR Technical Baseline [4] is shown in Exhibit 2-4. The inputs to the Production Model for the instruments and missions to be launched by June 2000 are currently derived from information provided to HAIS by the Ad Hoc Working Group for Production (AHWGP) and the Algorithm Theoretical Basis Documents (ATBDs) generated by each of the EOS instrument teams. The ATBDs describe the scientific rationale for each discrete product. The information from the ATBDs (and elsewhere) has been translated by the AHWGP into sets of science data production process characteristics for each instrument. Each EOSDIS epoch is subdivided by the quarter-year, into which the applicable processes are assigned. This information forms the basis for the

Performance Model's push workload characterization, over time: data arrival rates and volumes, process and archive physical location(s), process sequencing, inter-product dependencies, quality control, algorithm integration and test; and the computational, transient and archive data storage, and data transport requirements. These spreadsheet-based workload distributions are finally translated, manually, into the Performance Model's input parameter tables. As the ECS matures, it should be possible to calibrate the model with real-world information and generate workload predictions with progressively improving confidence.

The Performance Model is intended to provide a definitive basis for evaluating alternative ECS architectures capable of supporting user and production demands (as predicted by the User and Production Models), and to evaluate architectural sensitivities to predictive uncertainties. This model is implemented using the Block Oriented Network Simulator (BONeS) discrete-event simulation modeling tool. A top-level description of the current model under development is illustrated in Exhibit 2-5. As the model scope matures, subsystem models (currently, a mix of static and dynamic models) may be used to supply response characteristics. Doing so isolates their implementation details and mitigates the system-level model's execution resource demands, which tend to be extensive. The extent to which this may be done is still to be determined.

The Performance Model, when complete, should be capable of yielding several important categories of information by

- providing resource consumption statistics that could be used by user and production modeling personnel to assess the impact and improve the performance of their processes,
- identifying the driving parameters coupled with the capability of evaluating architectural sensitivities to their values in order to help focus analyses at minimizing their uncertainty,
- assessing expected performance requirements compliance by producing response-time and other statistics that can be compared directly to the requirements,
- establishing a firm basis for a bill of materials (BOM) and a set of operational requirements, such as media handling, necessary to implement, maintain, and operate the ECS for each epoch under consideration, and
- supporting the assessment of performance vs. cost impacts for new technologies that are being considered for inclusion into the ECS.

EXHIBIT 2-5: Major Components for the BONeS Performance Model

As the ECS matures, it should be possible to calibrate the model with real-world information and generate performance predictions with progressively improving confidence.

The Cost Model is intended to estimate the resources required to develop and operate ECS architecture alternatives (which can be partially derived from the Performance Model) within schedule constraints. The Cost Model is currently implemented as a collection of three types of stand-alone cost estimation models: Custom Software, Operations and Maintenance, and Commercial-Off-The-Shelf (COTS) Hardware and Software. There is no integrated Lifecycle Cost Model. Each model type may itself be composed of several different models or small variations of the same model. The transfer and correlation of information between the models and the aggregation of costs to calculate the overall lifecycle cost are largely manual operations. The IV&V team's ability to examine the details of these models has been very limited. If the models perform as they have been represented to us, then it should be possible to calibrate the models using actual ECS experience and generate future cost predictions with progressively improving confidence as the ECS matures.

This IV&V assessment was performed under EOSDIS IV&V Task 5B (Requirements Analysis & Traceability), specifically as part of Subtask 5.3 (ECS Modeling Assessment).

This page intentionally left blank.

3. USER MODEL

The User Model (i.e., pull model) is a representation of the interaction of Earth-science researchers and other potential users with the EOSDIS. It is intended to characterize a broad user profile describing *who* the users may be (demographics); *what* information they may need (products); *when,* by what means, and how often they are likely to access the system; and *how* they will use the system (services invoked). Results from the User Model (as well as from the Production Model in section 4) formulate the basis for determining the inputs (*how used*) to the Performance Model, and ultimately will provide information for the ECS system developers.

The findings of our independent assessment of the HAIS user-modeling activities include analysis tasks we performed; constraints affecting our analysis; and analysis results, conclusions, and recommendations, and are presented in this section.

3.1 Analysis Tasks Performed

3.1.1 Background

There are three major sources of load on the ECS: data push, user pull, and Version 0 migration. User pull consists of ECS users requesting services from the ECS. These services include data ingest, data subsetting, and data ordering. The time of occurrence of and the processor and data volume loads from user pull requests are not predictable (except in the stochastic sense); this is in contrast to data push and V0 migration, which are more or less regular and can be scheduled. Since the user pull demands on the ECS may be considerably larger than the other scheduled demands, it is important to have as reliable as possible user pull baseline on which to base stochastic inputs to the Performance Model. It is the adequacy and reliability of this user pull baseline that we are addressing in our analysis.

3.1.2 *Who* Are the Users (Demographics)

Who are the potential users of the EOSDIS, and what are their research interests? We have focused our analysis on the potential EOSDIS Earth-science users. We have analyzed part of the preliminary (therefore incomplete) version of the HAIS-produced EOSDIS and General Science User Survey (EGSUS) database, and compared these results with earlier HAIS demographic studies and with our independent demographic analysis [2].

3.1.3 *What* Are They Interested In (Products)

In order to estimate the potential loads (both processor and volume), it is necessary to determine which EOSDIS products might be requested by potential pull users. We have begun an analysis of the product requests contained in the preliminary EGSUS database and compared those findings to product requests contained in the HAIS'-produced user scenarios.

3.1.4 *How* Will They Use the System (Services Invoked)

We have begun a preliminary analysis of the services and functions invoked by the respondents to the EGSUS, and have compared them with earlier results from the user scenario database.

3.2 Constraints Affecting The Analysis

We did not receive the preliminary EGSUS databases from HAIS until June 26, 1995. The CDR Technical Baseline [4–in particular, *Attachment H. User Pull Baseline*)] was not made available to the public via the EDHS WWW server until June 29. In addition, we are not aware of any new formal white papers or technical reports from HAIS since we prepared our last TAR [3]. Most of the new results in this report are based on informal agendas of and notes from our biweekly meetings with members of the ESDIS Project and the HAIS modeling teams. Moreover, much of that work is still in progress. Therefore, all the new results in this section of this report must be considered preliminary—a report of work in progress.

3.3 Results

We have examined HAIS' modeling goal of creating user scenarios to make it possible to generate inputs to the ECS Performance Model. Starting with HAIS' science-user demography and proceeding through the user scenarios augmented by the EGSUS results, can they logically create the necessary inputs to the BONeS Performance Model that will be representative of the user requirements and the system requirements? Do the existing user scenarios and EGSUS-derived information contain sufficient information to accurately estimate the pull load on the system?

3.3.1 Discussion of Results

User Model: Top Level

Model Element	Model Characteristic	Analysis Priority	Level of Maturity	Analysis Date
1. User Profile (Who)	1.1 HAIS Demographics	1	2	31-Jul-95
	1.2 AHWGC Demographics	1	0	31-Jul-95
	1.3 EGSUS Demographics	1	2	31-Jul-95
2. User Needs (What)	2.1 EOS Products	1	1	31-Jul-95
	2.2 Other Products	2		
	2.3 IDS Products	2		
3. User Access (When)	3.1 In User Scenarios	1	1	10-Feb-95
	3.2 In AHWGC	1		
	3.3 In EGSUS	1		
	3.4 IDS-Required Access	2		
	3.5 V0 Access Statistics	2		
4. User Services (How)	4.1 In User Scenarios	1	1	31-Jul-95
	4.2 In AHWGC	1		
	4.3 In EGSUS	1	1	31-Jul-95
	4.4 IDS-Required Services	2		
5. Output to Performance Model (How Used)	5.1 Needs	1		
	5.2 Access	1		
	5.3 Services	1		

EXHIBIT 3-1: User Model Maturity Table

Exhibit 3-1 presents a summary of our results to date. We will address elements 1, 2, and 4 in this report. For those characteristics having priority 1 with no listed level of maturity, we do not have sufficient information available to warrant an analysis. These items will be analyzed as soon as the information becomes available to us.

3.3.1.1 User Profile (Who)

EGSUS Demography. HAIS converted the responses to this survey into an EGSUS database. The preliminary database contains 288 reliable records selected from the more than 500 responses to the survey. If, for instance, a respondent indicated a desire for global Landsat images for desertification studies, that response was rejected as being inappropriate and unreliable—perhaps an idle WWW browser (the survey was made available to the general public from an open WWW server). Each of the 288 reliable records amount to a user mini-scenario containing demographic information as well as product requests and requests for services. Each respondent was allowed to select one primary and as many as 5 secondary disciplines of interest. Exhibit 3-2 shows the distribution of these choices for the various disciplines listed in the survey. The left bars within each category represent the primary choices, and the right bars represent all choices (primary and all secondary).

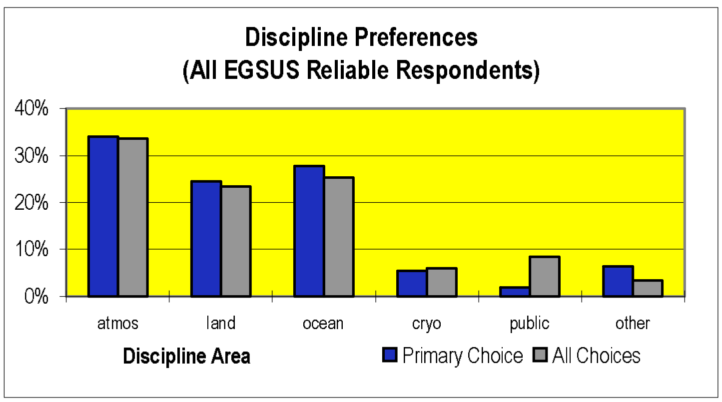


EXHIBIT 3-2: Discipline choices from the 288 reliable records

This distribution is to be compared to similar distributions, shown in Exhibit 3-3, from previous work. The left panel is from information contained in [5] and the right panel is from [3]. The relative distributions for the three major disciplines (atmosphere, land, and ocean) are decidedly different, and the distribution obtained from the user scenarios shows the land discipline to be the greatest. We will return to this troublesome point later in this section.

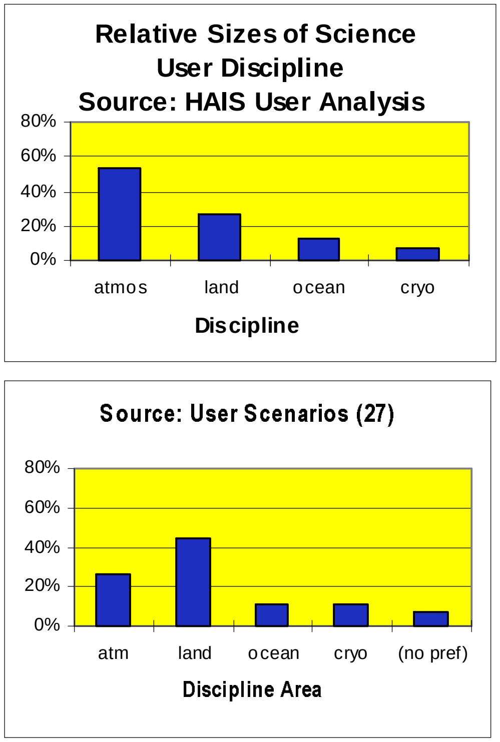


EXHIBIT 3-3: Discipline choices from previous work

In [2,3], we analyzed an independent database (a NASA Headquarters survey) for USGCRP research area preferences. That database contains 9,979 records. There are 127 users common to that database and the EGSUS database. The discipline areas (from the EGSUS database) for these selected users are shown in Exhibit 3-4. There are striking similarities between Exhibits 3-2 and 3-4 and equally striking dissimilarities among these two distributions and those of Exhibit 3-3. This is not too surprising because the left panel of Exhibit 3-3 is representative of the broad science community taken as a whole, while exhibits 3-2 and 3-4 are representative of a narrow sample of this community—potential users who have provided user mini-scenarios. These distributions may be more representative of the EOSDIS science-user community, but more work needs to be done before this can be stated with certainty.

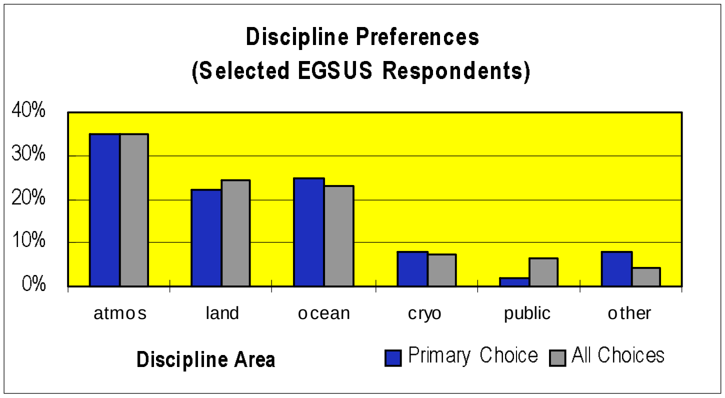


EXHIBIT 3-4: Discipline choices from the EGSUS Database

The USGCRP research area preferences (obtained from our NASA Headquarters database) for these common 127 records are shown in Exhibit 3-5. Besides reflecting the interdisciplinary nature of these potential EOSDIS users, this common link allows us to draw some comparisons between this group and previously analyzed groups.

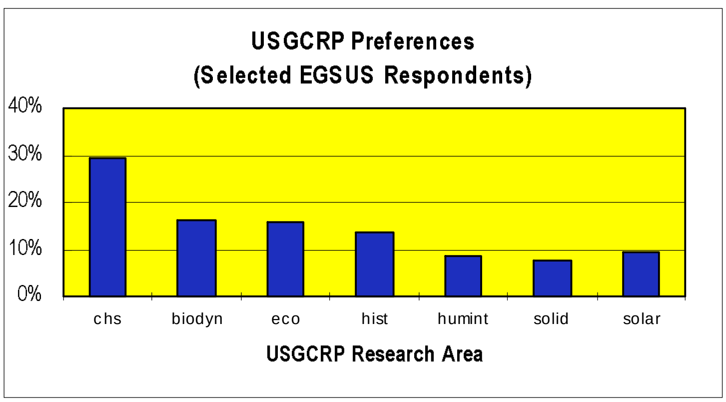
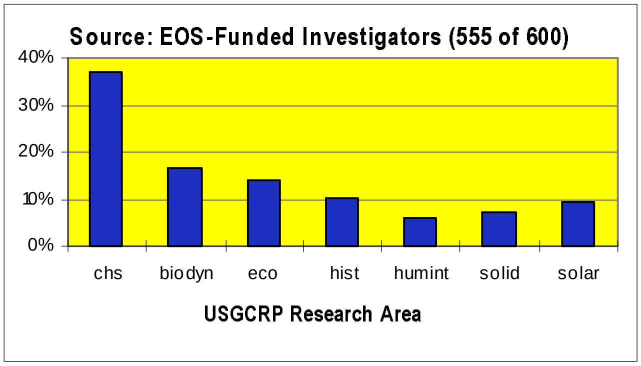


EXHIBIT 3-5: USGCRP Research Area Preferences Common to the EGSUS Database and the NASA HQ. Survey

Exhibit 3-6 shows two distributions of USGCRP research area preferences previously published in [3]. The left panel shows the preferences of 555 of the EOS-funded investigators that answered the NASA Headquarters survey. Their preferences are quite similar to those who were chosen from the EGSUS database. We have not yet verified whether the EGSUS database contains responses from some EOS-funded investigators. On the other hand, the right panel of Exhibit 3-6, thought at the time to be representative of potential EOSDIS science users, is decidedly different, suggesting that the filter we applied to the records in that database may not have been appropriate.



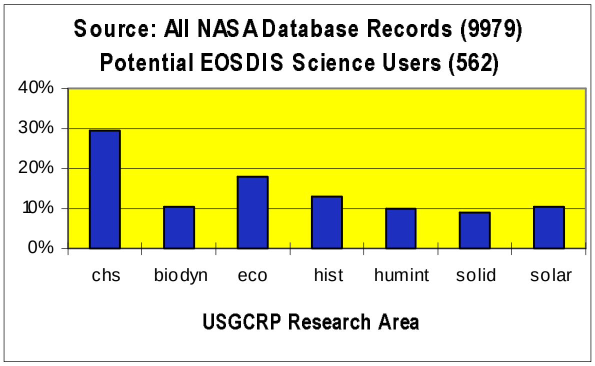


EXHIBIT 3-6: USGCRP Research Area Preferences from [3], for comparison

HAIS' User Scenario Matrix and the EGSUS. The user scenario matrix, Exhibit 3-7, is a HAIS-developed system for classifying potential EOSDIS science users according to their access methods and research data preferences. The six rows represent the anticipated method of accessing the system; the four columns represent their anticipated data needs. Twenty-seven potential users were interviewed by HAIS, the interviews were transformed into a user scenario database, and this database provides the necessary information (such as time and frequency of access, EOS products of interest and their sizes, and services invoked) for inputs to the Performance Model. We have placed into each cell the research discipline of the scenario provider. Forty-four percent of the scenarios represent land-oriented studies (see Exhibit 3-3). It is not clear whether or not this selection of users will bias the design of the system toward the research patterns and preferences of land-oriented users.

The EGSUS was designed so that the respondents could be placed into one of the elements of the six-by-four user scenario matrix by a decision tree based on their answers. In the past, the relative frequency of expected occurrence (or the probability that a random access to the system will fall into a cell of the matrix) for each of the 24 cells was determined solely from demographic information (such as that shown in

Exhibit 3-3, left panel). Use of the information from the EGSUS database may allow improvement of these probability estimates.

		Data Access Class			
		Reviews	Local/Field/ Case Studies	Regional Studies	Global Studies
System Access Class	Traditional User	varies	atmos	land	land
	Data Consumer	varies	land	land	land
	Data Browser	varies	land (1) atmos (1)	cryo (3)	atmos
	Analytical User	atmos	land	atmos	ocean
	Production User	n/a	land	ocean	land
	Machine-to-Machine	n/a	land (2)	atmos (2)	ocean

EXHIBIT 3-7: User Scenario Matrix, Showing Respondents Discipline Area

3.3.1.2 User Needs (What) and User Services (How)

Products. There are currently two sources of information for estimating which products EOSDIS science users will be interested in: from the user scenario database and from the EGSUS database. The user scenario database lists 110 unique (many were common to several scenarios) products. Not all of these products are listed in the CDR Technical Baseline [4], which contains 247 products (including four special on-request products from ASTER).

From the on-line (World Wide Web) version of the EGSUS:

All of the products to be produced by EOSDIS until the year 2000, and the DAO, LANDSAT-7, TMI, VIRS, PR, ERS, JERS, and RADARSAT products are included in the survey. Products from Version 0 and other heritage systems that will be migrated to EOSDIS, and archived and distributed from EOSDIS are not included in this survey.

The EGSUS database contains a list of 157 unique products from which the respondents may choose. Each of these products was selected by at least one of the 288 respondents. Some of these products do not appear in the SPSO product database (also contained in [4]) or the CDR Technical Baseline product lists. We have not yet completed a comparison of the EGSUS database product list with the CDR Technical Baseline (or SPSO) product list.

Services. There are 49 functions or service types (such as archive, ingest, search, manipulate, and exchange) requested in the user scenario database.[1,2] The EGSUS database contains two (browse and order) at the product-specific level and four (analysis, inspect, produce, and request) at the user-specific level. These last four are mainly of use in demographic estimates. And the demographic estimates are a strong point of the EGSUS. This is why earlier in this report we identified the 288 records in the EGSUS database as user mini-scenarios.

3.3.2 Identified Problems

It is still not clear that the user scenarios represent the expected distribution of Earth-science users of the ECS. The scenarios should be representative of the science-user interest so that the user service and data requirements can be correctly estimated and used as inputs to the Performance Model. Moreover, they do not request data from ADEOS II, RADAR ALT, and ACRIMSAT, which are expected to be available during 1998–2000. Even though it is indicated in the CDR Technical Baseline that the volume of data from ALT RADAR and ACRIMSAT is small, there should be some scenarios using data from these satellites as well as ADEOS II in order to identify additional functions or services that may be related to products from those sources. Whether the inclusion of data from EGSUS and the recently-initiated AHWGC survey will mitigate this shortcoming remains to be seen.

3.3.3 Potential Issues

In the User Scenarios, some users have identified several visualization utilities that would be advantageous for their research, and the corresponding functionality to be provided in the ECS is currently under study by HAIS. The results of this study were expected by PDR, and we are not aware that this has been done. The decisions on the visualization utilities should be expedited.

3.4 Conclusions

3.4.1 Technical Integrity

The user-modeling activities analyzed in this report can be grouped into two categories: user demographics (who are the users) and what do they want from the ECS (products and services). The information was taken from two primary sources: the user scenarios and the preliminary EGSUS database.

User Characterization and the User Scenarios. With the addition of the preliminary results from EGSUS, we feel that the distribution of expected user research area interests will be improved. Further, the possibility of assigning relative access probabilities to the cells of the user scenario matrix from the user mini-scenario information in the EGSUS database will improve the utility and reliability of this tool as a basis on which to build the necessary inputs to the Performance Model. This activity has now reached a level of somewhat limited maturity (characteristics 1.1 and 1.3 in Exhibit 3-1). The continuing EGSUS analysis and AHWGC survey analysis will provide additional necessary information.

Products and Services. The frequencies of occurrence of service invocations and product requests are needed inputs to the Performance Model. This information derived from the user scenarios is detailed (many service types) but sparse (27 scenarios). Similar information derived from the EGSUS database will contain fewer service types, but the product list is broader. Moreover, there are 288 user mini-scenarios in the preliminary database we analyzed. This is an area in which the AHWGC survey will play a crucial role, since the level of detail of the product and services requests is expected to be greater. This part of the user-modeling activities has reached a level of limited maturity (characteristics 2.1 and 4.1 in Exhibit 3-1).

3.4.2 User Satisfaction

The primary users of the work described and analyzed in this and previous documents are the HAIS system performance modelers and thence the system developers. We still cannot at this time determine whether the results of HAIS' user-modeling activities will be completely adequate to meet the performance-modeling and system-developing needs because much of the user-modeling work is still in progress. Nonetheless, the user-modeling work to date forms a good foundation for extending the work by incorporating our recommendations, and the improved model may provide sufficient and accurate enough information that can be used as input to the Performance Model without reservation.

3.4.3 Trends and Projections

The development of the HAIS User Model has improved substantially since our last TAR, mainly due to the gathering of additional information through the widely-available user survey, EGSUS. We expect that the AHWGC activities and survey will provide a similar improvement.

3.5 Recommendations

Based on the analysis conducted to date, we offer the following recommendations:

1. Expedite HAIS' analysis of the EGSUS and AHWGP survey.
2. Create and analyze as many additional scenarios as is necessary to be representative of the utilization of data from all EOS instruments available during the identified epochs, and truly reflect the international character of the science users by including scenario(s) from scientists or institution outside the United States.
3. Include functional or service requirements and pull load on the ECS that may arise from the access by the International science-user community—particularly International Partners (IPs).
4. Estimate the (increase in the) pull load on the ECS the due to non-EOS and non-satellite data after the expected increase in the scientific activity that will result from the general availability EOS data.

5. Estimate the (increase in the) pull load due to the access of satellite data from the International Partners (IPs), through ECS by the U.S. science users, and
6. Assign priorities to and binding of the user-generated requirements (from the user scenarios) to the available resources for each release.

3.5.1 Areas Requiring Further Analysis

Areas requiring further analysis are listed here.

1. Continue the analysis of the user characterization part of HAIS' User Model to further identify *who* the user is in terms of both USGCRP research areas and traditional Earth science disciplines based on the results we expect from the EGSUS and the AHWGC survey.
2. Verify that the changes to the User Model are made by HAIS and are based on the recommendations in this report.
3. Verify HAIS' methodology used to derive inputs to the Performance Model. Results from the EGSUS and AHWGC are expected to be important in this area. We recommend that the methodology used and the results obtained in the generation of inputs to the pull generator be subjected to verification as soon as possible, preferably before Release-B IDR.
4. Map the user scenario- and AHWGC-generated services and functions to the ECS Level 4 requirements.
5. Examine the adequacy of the User Model inputs (as they become available) to the BONeS Performance Model based on feedback from the results of performance modeling.

3.5.2 Solutions To Important Problems

1. The user characterization needs to be frozen (for the epochs under consideration) well before the Release-B IDR, based on the EGSUS-derived and AHWGC-derived updates to the model.
2. Complete the trade studies on user processing, visualization and browse products and take policy decisions where necessary early, so as to provide inputs into the model before the CDR.
3. Expedite evaluation of all functional or service requirements identified from the scenarios and the EGSUS and AHWGC surveys, identify the requested functions or services that cannot be implemented in the ECS, and evaluate the impact of omitting such functions or services.
4. The URDB process is a powerful tool for obtaining new user requirements. However, it is currently being used by a small number of potential investigators and is not representative of the broad user community in the United States. This limitation needs to be corrected by advertising the availability of this facility. Additionally, the URDB process needs to be expedited in order to be effective in providing information on new user requirements for the system modeling and its design.

3.5.3 Risk Management

The information that will be derived from the EGSUS and the AHWGC survey needs to be incorporated into the User Model as soon as possible in order to provide accurate and timely estimates of the pull load on the ECS by Earth-science users. The outputs from the User Model are used as inputs to the Performance Model which in turn drives the Cost Model. Therefore, accurate cost-modeling results will depend critically on accurate pull load modeling.

This page intentionally left blank.

4. PRODUCTION MODEL

The Production Model (i.e., data push model) is a representation of the science product generation processes (nominally, Level 0 input transformation to Levels 1/2/3/4 products). This model is intended to predict the steady-state and exceptional processing necessary to deliver trusted science data to the EOSDIS archives when required. Results from the Production Model formulate the basis for determining the inputs to the Performance Model and, ultimately, the ECS system developers.

The findings of our independent assessment of the HAIS production-modeling activities includes a list of analysis tasks performed; constraints affecting the analysis; and analysis results, conclusions, and recommendations, and are presented in this section.

4.1 Analysis Tasks Performed

4.1.1 Background

Of the three major sources of load on the ECS (data push, user pull, and V0 data migration), the data push load is perhaps the most straightforward to understand and analyze for a number of reasons, the most important being that this load is, for the most part, predictable. Therefore, it can be scheduled, in contrast to the user pull load, which must be handled stochastically.

The CDR Technical Baseline (TB) [4], prepared by the ECS Project, comprises the current state of knowledge of the push load—the EOSDIS products and their file sizes, processing requirements, frequency of and timelines for production, and so forth. The current TB contains these kinds of information from two sources. The Ad Hoc Working Group for Production (AHWGP) has prepared inputs for the TRMM and AM-1 missions, and the Landsat 7, ADEOS II (SeaWinds), and RADAR ALT missions. For the other missions, the SPSO database was used. The SPSO database reflects the updated mission baseline due to the restructuring activity. Exhibit 4-1 lists the files contained in the TB (see Appendix B for Attachment descriptions):

4.1.2 TB Structure and Methodology

To partially determine whether the TB contains the necessary information for the HAIS modelers and system developers to do their work, we analyzed the structure of selected parts of the TB and the methodology used in producing the input information necessary for the Production Model. In particular, we focused on the information contained in the Data Volume Summary, the AHWGP Processing Summary, the AHWGP Process and File Descriptions, the AHWGP Processing and Volume Timelines, the product Details, and the Parameter Details.

4.1.3 TB Self-Consistency

To partially determine whether the information in the TB is self-consistent, we examined some of the entries in the AHWGP process and file descriptions and processing and volume timelines spreadsheets, the product details spreadsheet, and the parameter details spreadsheet to see if they report the same information, and whether the information contained in the data volume and processing summary spreadsheets followed from these spreadsheets. We also partially examined the consistency between the AHWGP-and SPSO-based spreadsheets.

Technical Baseline Contents	
Technical Baseline (Main)	Text
Attachment A. Mission Baseline	Text
Attachment B. Data Volume Summary	Spreadsheet
Attachment C. Processing Summary	Spreadsheet
AHWGP—Tables Definitions	Text
Process Descriptions v2.1	Spreadsheet
File Descriptions v2.1	Spreadsheet
Processing Timelines v2.1	Spreadsheet
Volume Timelines v2.1	Spreadsheet
Attachment D. Product Details	Spreadsheet
Attachment E. Parameter Details	Spreadsheet
Attachment F. TSDIS Products	Text
Attachment G. V0 Migration Baseline	Text
Attachment H. User Pull Baseline	
Science Usage	Text
DAAC Pull Baseline	Spreadsheet
User Pull Info	Text
Attachment I. Requirements Baseline	Text
Attachment J. DAAC Implementation Baseline	
Schedule for years 1995–1998	Graphic
Schedule for years 1998–2001	Graphic
Attachment K. Capacity Phasing	Text
Attachment L. Summary of v2.1 Changes	Text

EXHIBIT 4-1: Contents of the CDR Technical Baseline

4.2 Constraints Affecting The Analysis

The CDR Technical Baseline was not made available to the public until June 27, 1995. Some of the work reported here, therefore, was begun using the PDR Technical Baseline. The details presented in this report are, however, from the CDR Technical Baseline.

4.3 Results

We have examined HAIS' production-modeling goal of using information in the TB to generate inputs to the BONeS Performance Model of the ECS. Is the information sufficient for performance-modeling (and thence system-design) purposes? Is it internally self-consistent?[3]

4.3.1 Discussion of Results

Exhibit 4-2 presents a summary of our results to date. We will address preliminary analyses of parts of elements 1 (AM-1) and 2 (TRMM) in this report. Because the CDR TB was just recently made available to us, we have not had sufficient time to analyze those characteristics having priority 1 with no listed level of maturity (MODIS and DAO).

Top Level — Production Model				
Model Element (Mission)	**Model Characteristic (Instrument or Product)**	**Analysis Priority**	**Level of Maturity**	**Analysis Date**
1. AM-1	1.1 ASTER	1	2	31-Jul-95
	1.2 CERES	1	1	31-Jul-95
	1.3 MISR	2		
	1.4 MODIS	1		
	1.5 MOPITT	2		
2. TRMM	2.1 CERES	1	1	31-Jul-95
	2.2 LIS	3		
	2.3 Other (VIRS, PR, TMI)	3		
3. Other (Missions & Products)	3.1 Landsat-7	2		
	3.2 Follow-on Missions	3		
	3.3 DAO Products	1		
	3.4 Version 0 Products	2		
	3.5 IDS-Generated Products	2		
	3.6 External Products	3		

EXHIBIT 4-2: Production Model Maturity Table

4.3.1.1 TB Structure and Methodology

The HAIS Performance Model is implemented through BONeS, a COTS, block-oriented network simulator. Each process modeled requires an input from the Production Model, such as input and output file size, processing requirements, and frequency of occurrence. This information is contained in the various spreadsheets making up the TB. We will step through this path for one process as an example of the methodology, examining the structure of the spreadsheets as we move among them.

The process description spreadsheet in the TB was prepared by HAIS from information provided by the AHWGP. It lists the processes that produce standard and other products for TRMM (CERES, LIS), AM-1 (ASTER, CERES, MISR, MODIS, MOPITT), METEOR (SAGE III), ADEOS (SeaWinds), and RADAR ALT (DFA, MR). It includes the input and output file names for each process, when the process will be executed (epoch), number of executions per day, and millions of floating point operations per execution.

The process we chose for analysis is the product generation executable (PGE) that produces the standard product AST10 (Scene Classification) from the ASTER instrument (chosen because it is the first entry in the spreadsheets). The first six rows of the part of the TB process description spreadsheet that is shown in Exhibit 4-3 contain the names of the six files needed (i.e., the data dependencies) by the process AST_PGE_01 in order to produce the output file AST_10 (AST10). This process is executed 182 times each day and uses 4,993 million floating point operations for a processing load of 10.52 MFLOPS. It is executed from epoch g (3Q 98) through epoch x (4Q 02).

Process ID	Process Name	Processing Site	Epochs	Input File IDs	# Read per Execution	Amount Read (Fraction)	Output File ID	# Written per Execution	Amt Wrtten (Fraction)	Millions of Floating Point Ops per Execution	No. of Exec. /day
AST_PGE_01	Scene classification (stand. processing)	EDC	ghijklmnopqr stuvwx	AST_ANC_01	1	1	AST_10	1	1	4,993	182.00
				AST_ANC_02	1	0.001					
				ANC_EDC_DEM	1	0.001					
				Anc_EDC_LANDCOVER	1	0.001					
				AST_L1B	1	1					
				ANC_DCW_Land/Sea	1	1					
AST_PGE_02	Decorrelation stretch	EDC	ghijklmnopqr stuvwx	AST_L1B	1	1	AST_06B	1	1	18,337	182.00
				AST_10	1	1	AST_06C	1	1		
				AST_DS_TMP1	1	1	AST_06A	1	1		
				AST_DS_TMP2	1	1					

EXHIBIT 4-3: From the AHWGP Process Description Details v2.1

The six input files and the sole output file for this process are described in the TB file description spreadsheet, shown partially in Exhibit 4-4. This, besides identifying the files by name, shows the file sizes, the SPSO name, the file location, and the file disposition. Note that one of the files (AST_L1B) is itself an ASTER product (Level 1B registered radiance at the sensor), and must be available before the AST_PGE_01 process is run. Returning to the TB process description spreadsheet (Exhibit 4-3), the fraction of each dependency file to be read is shown, and (together with the file size from Exhibit 4-4) provides the necessary information to determine the file I/O requirements for each execution of the process producing AST_10. This information, taken all together, provides the necessary information for the HAIS performance modelers to construct the input tables for the BONeS Performance Model, discussed in section 5 of this report. This methodology must be followed for every process to be modeled.

File ID	Instrument	File Name	SPSO Equivalent	Archive Site	File Disposition	File Size (MB)	Temporal Coverage (Minutes)	Root/External Flag	Ingest Media Flag	Source
AST_10	**ASTER**	**scene classification_3804**	**AST10 (3804)**	**EDC**	**Archive**	**18**	**7.91**	**0**	**0**	**2**
AST_ANC_01	ASTER	ASTER Classification Coefficient File_TBD	none avail	EDC	Permanent	0.5	0.00	0	0	2
AST_ANC_02	ASTER	ASTER Nominal Atmospheric Coefficient File_Gillesp	none avail	EDC	Permanent	1	0.00	0	0	2
ANC_EDC_DEM	Other	Digital Elevation Map	N0002	EDC	Permanent	200	500,000.00	0	0	2
ANC_EDC_LANDCOVER	Other	Surface Land Cover and Vegetation Type	N0006	EDC	Permanent	250	129,600.00	0	0	2
AST_L1B	ASTER	registered radiance at sensor_2452	AST03(2542)	EDC	Archive	243	7.91	1	1	11
ANC_DCW_Land/Sea	Other	Land/Sea Boundary Data	N0030	EDC	Permanent	80	500,000.00	0	0	2

EXHIBIT 4-4: From the AHWGP-Based TB File Description Details v2.1

The processing information is summarized in the TB processing summary spreadsheet, partly shown in Exhibit 4-5. Notice that the process AST_PGE_01 shows 10.52 MFLOPS, as calculated from the description files. Also, the process AST_PGE_09, producing the polar cloud map, is not scheduled for execution until epoch m (1Q 00).

Process ID	Process Name	Exec. Site	Epochs	MFPOs	#/day	MFLOPS	2Q 98 epoch f	3Q 98 epoch g	4Q 98 epoch h	1Q 99 epoch i	2Q 99 epoch j	3Q 99 epoch k	4Q 99 epoch l	1Q 00 epoch m
ASTER														
AST_PGE_01	Scene classification (stand. processing)	EDC	ghijklmnopqrstuvwx	4,993	182.00	10.52	0.00	10.52	10.52	10.52	10.52	10.52	10.52	10.52
AST_PGE_02	Decorrelation stretch	EDC	ghijklmnopqrstuvwx	18,337	182.00	38.63	0.00	38.63	38.63	38.63	38.63	38.63	38.63	38.63
AST_PGE_03	Brightness temperature	EDC	ghijklmnopqrstuvwx	861	70.00	0.70	0.00	0.70	0.70	0.70	0.70	0.70	0.70	0.70
AST_PGE_04	Atmospheric correction--VNIR, SWIR	EDC	ghijklmnopqrstuvwx	34,200	70.00	27.71	0.00	27.71	27.71	27.71	27.71	27.71	27.71	27.71
AST_PGE_05	Atmospheric correction--TIR	EDC	ghijklmnopqrstuvwx	5,221	70.00	4.23	0.00	4.23	4.23	4.23	4.23	4.23	4.23	4.23
AST_PGE_06	T/E	EDC	ghijklmnopqrstuvwx	3,329	70.00	2.70	0.00	2.70	2.70	2.70	2.70	2.70	2.70	2.70
AST_PGE_09	Polar cloud map	EDC	**mnopqrstuvwx**	102,600	6.00	7.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.13
AST_PGE_10	DEM	EDC	ghijklmnopqrstuvwx	513,000	1.00	5.94	0.00	5.94	5.94	5.94	5.94	5.94	5.94	5.94
ASTER							**0.00**	**90.41**	**90.41**	**90.41**	**90.41**	**90.41**	**90.41**	**97.54**

EXHIBIT 4-5: From the AHWGP-Based TB Processing Timelines v2.1

The data volume information is summarized in the AHWGP-based TB volume timelines spreadsheet, part of which is shown in Exhibit 4-6. Note that the file AST_10 produced by the process AST_PGE_01 does not appear in this spreadsheet. We will return to this point in the section immediately following.

File ID	File Name	Instrument	Archive Site	Product Size (In MBs)	# Prod. Files Written/Execution	# Times Prod. Process Executed/Day	Daily Volume (GB)	Epochs	3Q 98 (epoch g)	4Q 98 (epoch h)	1Q 99 (epoch i)	2Q 99 (epoch j)	3Q 99 (epoch k)	4Q 99 (epoch l)	1Q 00 (epoch m)
ASTER															
AST_L1A	Lev 1A from Japan	ASTER	EDC	125	1	182	22.75	ghijklmnopqrstuvwx	22.75	22.75	22.75	22.75	22.75	22.75	22.75
AST_L1B	registered radiance at sensor_2452	ASTER	EDC	243	1	182	44.23	ghijklmnopqrstuvwx	44.23	44.23	44.23	44.23	44.23	44.23	44.23
AST_06A	Decorrelation Stretch - VNIR	ASTER	EDC	52.8	1	182	9.61	ghijklmnopqrstuvwx	9.61	9.61	9.61	9.61	9.61	9.61	9.61
AST_06B	Decorrelation Stretch - SWIR	ASTER	EDC	13.2	1	182	2.40	ghijklmnopqrstuvwx	2.40	2.40	2.40	2.40	2.40	2.40	2.40
AST_06C	Decorrelation Stretch - TIR	ASTER	EDC	1.1	1	182	0.20	ghijklmnopqrstuvwx	0.20	0.20	0.20	0.20	0.20	0.20	0.20
AST_04	Brightness Temperature At Sensor	ASTER	EDC	6	1	70	0.42	ghijklmnopqrstuvwx	0.42	0.42	0.42	0.42	0.42	0.42	0.42
AST_09A	Surface radiance--VNIR	ASTER	EDC	176	1	70	12.32	ghijklmnopqrstuvwx	12.31	12.31	12.31	12.31	12.31	12.31	12.31
AST_09B	Surface radiance--SWIR	ASTER	EDC	61.6	1	70	4.31	ghijklmnopqrstuvwx	4.31	4.31	4.31	4.31	4.31	4.31	4.31
AST_09C	Surface radiance--TIR	ASTER	EDC	6	1	70	0.42	ghijklmnopqrstuvwx	0.42	0.42	0.42	0.42	0.42	0.42	0.42
AST_07A	Surface reflectance VNIR	ASTER	EDC	176	1	70	12.32	ghijklmnopqrstuvwx	12.32	12.32	12.32	12.32	12.32	12.32	12.32
AST_07B	Surface reflectance SWIR	ASTER	EDC	62	1	70	4.34	ghijklmnopqrstuvwx	4.34	4.34	4.34	4.34	4.34	4.34	4.34
AST_05	**Surface Emissivity**	**ASTER**	**EDC**	**5**	1	**70**	**0.35**	**ghijklmnopqrstuvwx**	0.35	0.35	0.35	0.35	0.35	0.35	0.35
AST_08	**Surface Kinetic Temperature**	**ASTER**	**EDC**	**1**	1	**70**	**0.07**	**ghijklmnopqrstuvwx**	0.07	0.07	0.07	0.07	0.07	0.07	0.07
AST_13	Polar Cloud Mask	ASTER	EDC	18	1	6	0.11	**mnopqrstuvwx**							0.11
AST_14	ASTER DEM Product	ASTER	EDC	35	1	1	0.04	ghijklmnopqrstuvwx	0.04	0.04	0.04	0.04	0.04	0.04	0.04
ASTER									**113.78**	**113.78**	**113.78**	**113.78**	**113.78**	**113.78**	**113.88**

EXHIBIT 4-6: From the AHWGP-Based TB Volume Timelines v2.1

Data Volume for At-Launch and Post-Launch Data Products (As Modified by AHWGP Results)											
					Version 1.2 (Prepared 7/22/94/Modified 11/8/94 ,12/21/94 & 6/21/95)						
Platform	Launch		Avg. Data Rate		Daily Data Volume (GB/day)				Total *	PDR	De
	Date	Instrument	(Kbps)	L-0 **	L-1A	L-1B	L-2	L-3/4	(L0 - L4)	Baseline	
		CERES	10.000	0.090	Use AHWGP Inputs (Totals show peak values)				12.310	11.350	
TRMM	Aug-1997	LIS	6.000	0.065	Use AHWGP Inputs (Totals show peak values)				1.620	1.330	
		Total	16.000	0.155					13.930	12.680	
		ASTER	8300.000	89.640	Use AHWGP Inputs (Totals show peak values)				113.880	118.810	
		CERES	20.000	0.181	Use AHWGP Inputs (Totals show peak values)				16.230	11.900	
AM-1	Jun-1998	MISR	3800.000	41.040	Use AHWGP Inputs (Totals show peak values)				148.450	136.390	
		MODIS	6200.000	66.960	Use AHWGP Inputs (Totals show peak values)				518.000	633.730	-
		MOPITT	6.000	0.065	Use AHWGP Inputs (Totals show peak values)				0.170	0.190	
		Total	18326.000	197.886					796.730	901.020	-
FOO	May-1998	COLOR	347.220	3.750	0.000	5.500	0.461	0.536	10.247	10.247	
ADEOS II	Feb-1999	SWS	5.100	0.055	Use AHWGP Inputs (Totals show peak values)				5.050	0.557	
ALT RADAR	Mar-1999	DFA	1.375	0.015	Use AHWGP Inputs (Totals show peak values)				0.330	0.171	
		MR	0.125	0.001	Use AHWGP Inputs (Totals show peak values)				0.002	0.001	
		Total	1.500	0.016	0.000	0.000	0.000	0.000	0.332	0.172	
ACRIMSAT	Jun-1999	ACRIM	1.000	0.011	0.046	0.000	0.000	0.000	0.057	0.057	
RSA/CNES	Jan-2000	SAGE III	24.267	0.262	0.000	0.016	0.002	0.000	0.280	0.280	
SPACESTATION	Jun-2000	SAGE III	24.267	0.262	0.000	0.016	0.002	0.000	0.280	0.280	
		AIRS	1420.000	15.336	16.200	14.682	1.548	0.000	47.766	47.766	
		AMSU-A	3.200	0.035	0.033	0.033	0.000	0.000	0.101	0.101	
PM-1	Dec-2000	CERES	20.000	0.181	Use AHWGP Inputs (Totals show peak values)				9.100	12.220	
		MHS	4.200	0.045	0.042	0.042	0.000	0.000	0.129	0.129	
		MIMR	67.000	0.724	0.000	5.100	0.069	0.013	5.906	5.906	
		MODIS	6200.000	66.960	175.000	519.200	173.104	3.988	938.252	938.252	
		Total	7714.400	83.280	191.275	539.057	174.721	4.001	1001.254	1004.374	
		HIRDLS	40.000	0.432	0.000	0.634	0.095	0.090	1.251	1.251	
CHEM	Dec-2002	MLS	5.000	0.054	0.200	0.200	0.200	0.015	0.669	0.669	
		TES	3240.000	34.992	0.000	30.000	0.501	0.000	65.493	65.493	
		Total	3285.000	35.478	0.200	30.834	0.797	0.105	67.413	67.413	
ALT LASER	Jul-2003	GLAS	100.000	1.080	8.119	0.500	0.147	0.000	9.846	9.846	
Note:											
1. Data volume estimates are for at-launch and post-launch data products, excluding interim, special, and on-request products.										Total Delta	
2. * Totals for AHWGP instruments do not include Level 0 data											
3. ** Level 0 data for ASTER is not archived by ECS, L0 delievered to Japan by EDOS											
4. For CERES data volume specified reflects additional capacity required beyond previous mission											

EXHIBIT 4-7: CDR Data Volume Summary Spreadsheet

We need to point out that these processing and volume timelines spreadsheets do not themselves contain the level-of-detail information necessary to construct the inputs for the BONeS Performance Model. Those inputs must come from the TB processing description and file description spreadsheets, because this is where the needed file dependencies are presented for each process that is modeled.

Finally, a summary of the processing and volume information from the processing timelines and volume timelines spreadsheets is presented in the data volume summary and processing summary spreadsheets. Exhibit 4-7 shows the data volume summary. (The corresponding spreadsheet for the processing summary is not shown here.)

4.3.1.2 TB Self-Consistency

Focusing on ASTER as an example, we examined all the relevant spreadsheets in the TB for self-consistency (Are the corresponding entries the same?) and completeness (Is each file, process, or product accounted for? Are any missing from the spreadsheets where they are needed?).

Processing Loads. Exhibit 4-8 depicts the eight processes (PGEs) and fourteen output products represented in the process description spreadsheet for ASTER and summarizes the corresponding processing load. The processing load shown in the last column agrees with the corresponding entries in the processing timelines spreadsheet, and the total (which represents the peak load) agrees with the corresponding entry in the processing summary spreadsheet. Moreover, the peak processing loads represented in the processing summary spreadsheet for the other instruments for which the AHWGP provided detailed information to HAIS agree with the corresponding information in the processing timelines spreadsheet, with the exception of those for CERES, to which we will return later.

Process	Products	MFPOs	# of times	MFLOPS
AST_PGE_01	AST_10	4,993	182	10.52
AST_PGE_02	AST_06A, AST_06B, AST_06C	18,337	182	38.63
AST_PGE_03	AST_04	861	70	.70
AST_PGE_04	AST_07A, AST_07B, AST_09A, AST_09B	34,200	70	27.71
AST_PGE_05	AST_09C	5,221	70	4.23
AST_PGE_06	AST_05, AST_08	3,329	70	2.70
AST_PGE_09	AST_13	102,600	6	7.13
AST_PGE_10	AST_14	513,000	1	5.94
8	**14**	**—**	**—**	**97.54**

EXHIBIT 4-8: Aster Processing Load; Bottom Row Shows Totals

Data Volume Loads. The data loads represented in the file description should agree with those in the volume timelines spreadsheet. However, two files (AST_L1A and AST_L1B) represented in this spreadsheet do not appear in the process description spreadsheet (Level 0–Level 1 production is done in Japan), and one file (AST_10—produced by the process AST_PGE_01, shaded in Exhibit 4-8) that is listed in the process description spreadsheet does not appear in the volume timeline spreadsheet. This file is listed as being archived and produced 182 times each day for a total of 3.276 GB/day. However, this product (Scene Classification) is no longer a standard product, and will be archived for a period of TBD months, according to the TB summary of changes. The ECS performance modelers are looking at making this an archivable product with a six-month rolling storage life.

ADEOS II and ALT RADAR—A Random Spot Check. We briefly examined the ADEOS II (SeaWinds) and ALT RADAR (DFA) entries in the various TB spreadsheets for self-consistency. The results are shown in Exhibit 4-9.

CDR Source Spreadsheet	Vol. (GB/day)	Proc. (MFLOPS)	Vol. (GB/day)	Proc. (MFLOPS)

	ADEOS II SWS	ADEOS II SWS	ALT RADAR DFA	ALT RADAR DFA
Data Volume Summary	5.05	n/a	0.330	na
Processing Summary	n/a	0.017	NA	0.000
Volume Timelines	3.30	n/a	0.3306	na
Processing Timelines	n/a	0.017	na	0.001
Product Details	0.502	0.878	0.156	4.5

EXHIBIT 4-9: ADEOS II (SWS) and RADAR ALT (DFA) Summary

There are differences, showing a lack of self-consistency. In particular, the 4.5 MFLOPS figure in the last row, last column appears to be the processing load for ALT RADAR DFA instrument from the PDR Technical Baseline.

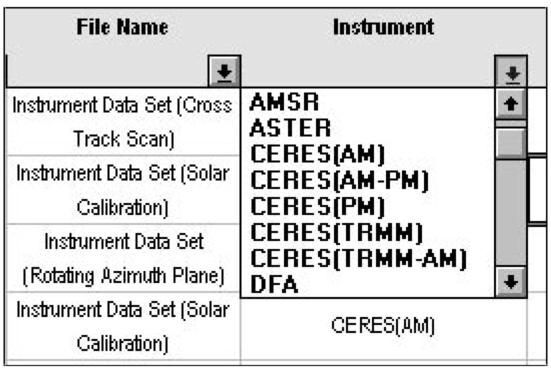


EXHIBIT 4-10: CERES Choices from the Volume Timelines Spreadsheet

CERES—A Quick Look. The CDR Mission Baseline lists the CERES instrument on three missions: TRMM, AM-1, and PM-1. Our attempts to examine the processing and data volume self-consistency led to a possible inconsistency in the volume timelines spreadsheet. There are five choices available for sorting the spreadsheet by CERES instruments (see Exhibit 4-10, which is a reproduction of the drop-down selection box for sorting the entries in the spreadsheet by instrument name). We could not, in the brief time we have had the CDR TB available for analysis, find any combination of sorting the instrument-mission combinations that would lead to the data volume loads presented in the data volume summary spreadsheet. We could not reproduce the peak processing load that is entered in the processing summary spreadsheets either.

4.3.2 Identified Problems

The structure of the current version of the CDR TB contains the information necessary for the HAIS performance modelers to derive the input data for their Performance Model. Lack of complete self-consistency among the various spreadsheets contained in this TB, however, may lead to spurious Performance Model outputs.

4.3.3 Potential Issues

If the CDR TB, and any subsequent version of the TB, is to be used for ECS performance modeling and ECS system design and development, then it is necessary that they be error-free, complete, and self-consistent. These TBs are the keystone for ECS system design. The matter of archiving such products as AST10 (scene classification) as well as the processing and distribution of special on-demand products, such as AST15–18, may affect both the modeling and design of the ECS.

4.4 Conclusions

4.4.1 Technical Integrity

The production-modeling–related activities analyzed in this report can be grouped into two categories: the translation of TB-contained information into suitable Performance Model inputs (the structure and methodology) and a preliminary assessment of the self-consistency of the information in the TB. *Methodology*. The process by which the TB information contained in the process description and file description spreadsheets can be converted into the required Performance Model input tables is sound and expected to be reasonably complete. This is mainly a result of the cooperative work between HAIS and the AHWGP in providing the level-of-detail information that was used to construct these process description and file description spreadsheets. Modeling instrument processes for the instruments not having this detailed information will be difficult. This part of the Production Model has achieved a level of somewhat limited maturity for ASTER and limited maturity for CERES (see Exhibit 4-2).

Self-consistency. The outputs from the production-modeling activities to the Performance Model need to be completely accurate and self-consistent in order that the outputs from the Performance Model to both the Cost Model and the ECS system developers be correct. Accuracy is a matter between the Instrument Teams, but self-consistency is a HAIS matter. Our preliminary analysis indicates that there are problems in the latter area. For this reason, this part of the HAIS production-modeling activities have achieved a level of somewhat limited maturity for ASTER and limited maturity for CERES (see Exhibit 4-2).

4.4.2 User Satisfaction

The primary users of the work described and analyzed in this and previous documents are the HAIS system performance modelers and thence the system developers. We expect at this time that the results of HAIS' production-modeling activities—especially the TB—will eventually be adequate to meet the performance-modeling and system-developing needs. The production-modeling work to date forms a good foundation for extending the work by incorporating our recommendations, and the improved model may provide the necessary, sufficient, and accurate information needed as input to the Performance Model without reservation. The methodology is sound; the execution of the details may be flawed.

4.4.3 Trends and Projections

The purpose of this section is to highlight measurable differences observed between the results of the current analysis and the previous one; and to project the implications of those differences into the future (i.e., whether they appear to be diverging from, or converging toward a reliable Production Model. The methodology has improved, and it can be expected to improve further should our recommendations be followed. The self-consistency has not improved, but this area, too, can be improved by following our recommendations.

4.5 Recommendations

4.5.1 Areas Requiring Further Analysis

The breadth of IV&V Production Model analysis needs to be expanded to include all missions and instruments that play a significant role in the production loading of the ECS. Now that analysis input data is becoming more readily available, this effort should prove to be more illuminating in terms of maturity level quantification. Future IV&V analyses (prioritized by workload impact) can be expected to yield maturity metrics by mission, instrument, and process, together with accompanying engineering analysis rationales.

4.5.2 Solutions To Important Problems

We recommend the following:

1. Ensure that all entries in all spreadsheets in the CDR and future Technical Baselines are complete, self-consistent, and accurate. In particular, the information contained in the product details and parameter details spreadsheets (from the SPSO product and parameter databases) needs to be current and should reflect current information in the AHWGP-information–derived spreadsheets. It is confusing to have incorrect, inconsistent information in these databases for the AHWGP instruments. Even though the TB introduction[4] states, “The information in this file should be used for all the products of all instruments not included in the AHWGP information. ...but where there is any conflict, the AHWGP information has precedence,” the possible propagation of incorrect information may have serious and unexpected consequences.
2. For the instruments and processes not already considered by the AHWGP, construct within the process and file description the same type and level-of-detail information as is already there for the AHWGP-based processes. Without this level-of-detail information, it will be very difficult to model these processes for those other instruments. This will require input from the Instrument Teams of the remaining instruments. The same standards of completeness, self-consistency, and accuracy should apply to these updated data.
3. Modify the format of the CERES information contained in the volume timelines spreadsheet (and wherever else it occurs) so that it is easier to clearly and unambiguously identify projected system loads according to what

platform the instrument is on (TRMM, AM-1, or PM-1). This is an ease-of-use modification, and will enhance the usability of the data.

4. Expand the level of information in the V0 Migration Baseline in the TB. It currently provides only gross estimates. Although V0 data migration is not a formal part of the data push modeling, it is being modeled by the HAIS performance-modeling team, and the process will compete for system resources and must be scheduled optimally. A database (spreadsheet) should be prepared indicating the individual data products expected to be migrated together with their file sizes and timelines.

4.5.3 Risk Management

An important risk factor in the development of the ECS is the inclusion of all relevant information about the data push load. This development must avail itself of all information from any modeling activities so that design problems can be identified as early as possible. Data push modeling should be accelerated, and steps need to be taken that will ensure the accuracy and reliability of the modeling activities.

This page intentionally left blank.

5. PERFORMANCE MODEL

The Performance Model is intended to provide a basis for evaluating the performance of an ECS design capable of supporting user and production demands (as predicted by the User and Production Models), and evaluating design sensitivities to predictive uncertainties and evolving requirements. This model is implemented using the **B**lock **O**riented **Ne**twork **S**imulator (BONeS) discrete-event simulation modeling tool. The findings of this independent assessment of the performance modeling activities, including analysis tasks performed, constraints affecting the analysis, analysis results, conclusions, and recommendations, are presented in this section.

5.1 Analysis Tasks Performed

5.1.1 Evaluation Areas

The purpose of this analysis is to evaluate the HAIS ECS Performance Model with focus on the detailed design of Release A SCDO. The analysis areas and the evaluation methodology are essentially the same as documented in our February Modeling Assessment TAR [3]. Specifically, the evaluations performed are as follows:

- Evaluate representation of the system design
- Evaluate representation of the system workloads
- Evaluate statistics collection/performance metrics quantification
- Evaluate overall model structure

5.1.1.1 System Design Representation Evaluation

The first evaluation area involves assessing the BONeS model representation of the ECS system design, specifically: the SDPS subsystems; the CSMS networks; and the distributed system resources.

Our analysis approach is to examine the BONeS model modules and parameters for completeness and correctness for the subsystem functional definitions of SDPS and CSMS, the resources of the distributed system design (i.e., processors, internal I/O channels, storage devices, local area networks, and wide area networks), and the internal and external interfaces of ECS. The nature of these model elements and the rationale for inclusion of this evaluation area are described in the previous version of this TAR [3].

5.1.1.2 Workloads Evaluation

In the second evaluation area, we evaluate the inclusion of ECS workload requirements in the BONeS model. These workload areas evaluated are:

- Push
- Pull
- Reprocessing
- V0
- System Overheads

Our analysis approach is to examine the BONeS model modules and parameters for completeness and correctness with respect to workload frequencies and flows through the system resources. The system overhead category is included to account for system performance degradation due to system functions such as operating systems, database and file management systems, and communication protocols. The nature of these model elements and the rationale for inclusion of this evaluation area are described in the previous version of this TAR [3].

5.1.1.3 Performance Statistics Evaluation

The third evaluation area involves assessing performance statistics collection, model results, and model validation.

Our analysis approach is to: 1) examine the statistical probe modules and the simulation configuration file in the BONeS model to determine what performance statistics are collected by the model for assessing system resource utilization and performance requirements, 2) evaluate the statistical soundness of the data analysis methods for simulation results; 3) evaluate the validation performed by HAIS on the performance predictions produced by the model, and 4) evaluate the adequacy of performance studies conducted with the model. The nature of the statistical collection model elements and the rationale for inclusion of this evaluation area are described in the previous version of this TAR [3].

5.1.1.4 Model Structure Evaluation

In the fourth evaluation area we analyze the overall model structure for the purpose of determining how easy it would be to modify the model to address deficiencies, add additional levels of detail or functionality, or make architectural changes.

Our evaluation approach is derived from the three previously listed evaluations. We examine the model's implementation of system functions, design, workloads, and statistical data collection in terms of modules, data structures, and parameters. The nature of the model structural evaluation and the rationale for inclusion of this evaluation area are described in the previous version of this TAR [3].

5.1.2 Evaluation Metrics

The evaluation metrics used for the Performance Model assessment are categorized as follows:

- Engineering Quality
- Testability
- Traceability
- Flexibility

5.1.2.1 Engineering Quality Metric

The Engineering Quality metric is defined in terms of the completeness and correctness of the model in representing the ECS subsystem functions, the distributed system design, and the system users. The quality of the model's results are directly affected by the completeness and correctness of the system representation. The model should include all sources of instrument product workloads and all user service requests. It should include all the subsystems and the distributed computer and communication resources of the design.

This metric is partially a measure of ECS functional traceability since it measures the completeness of the representation of the ECS functions and workloads. Further, it measures the accuracy of the design representation and its associated performance parameters such that components can be sized and performance requirements can be traced (see Traceability metric below).

5.1.2.2 Testability Metric

The Testability evaluation metric is defined in terms of the methods used for validating model results. Performance metrics should be defined for the model such that corresponding empirical measurements can eventually be made. Short of system testing, the model results can be validated by analysis techniques. For example, queueing theory and statistical analysis techniques can be used to validate that the simulation results are reasonable and to determine bounds on performance metrics.

5.1.2.3 Traceability Metric

The Traceability metric is defined in terms of the ability of the model to assess design performance and track performance requirements. The statistics collected in the model should provide the data to make these assessments. The computer and communication resources have to be sized so that the performance requirements (e.g., response time) can be achieved. It is essential not only to associate the performance requirements to the model components but also to determine what performance drivers impact requirements compliance.

5.1.2.4 Flexibility Metric

The Flexibility evaluation metric measures the model’s ability to evolve. The model may need to be modified to evaluate different releases of the system, additional levels of detail, performance requirements compliance, and changes in architecture and design. It may also need to be modified to address changes (proposed or actual) in requirements, new users , more (or less) frequent use, or different scenarios of use (i.e., evolution in how the users use the system).

The flexibility of the model to accommodate changes is probably the major factor in determining its long term usefulness as a performance analysis tool for the ECS program.

5.2 Constraints Affecting The Analysis

The primary constraint affecting the analysis of the Performance Model is limited simulation results. The results available focus primarily on ingest and production of standard products for TRMM and AM-1 instrument loads.

A secondary constraint is the lack of CDR-level documentation for the SCDO detailed design. Our assessment of the model's representation of the design is based on SDPS PDR documentation and other information gathered through informal channels with HAIS engineers since PDR.

5.3 Results

5.3.1 Discussion of Results

The results of the Performance Model evaluation are discussed below and are organized by four evaluation areas: system representation; workloads; performance statistics; and model structure. The model evaluation is based on the following sources: the SDPS Segment Design Specification from PDR [7]; discussions, briefings, and a performance modeling white paper [6]; various performance modeling handouts from HAIS; and a recent version of the BONeS Performance Model (Version 6z) with input parameter files for epoch e (TRMM) and epoch k (TRMM and AM-1). Other documentation from the pre-PDR timeframe was also examined and is listed in the previous version of the TAR [3]. Results of our evaluation are summarized in Exhibit 5-1. As shown in the exhibit, the four evaluation areas are broken into multiple categories. Each of the categories is given an analysis priority, level of maturity, and analysis date. Results reported reflect the state of the model as of 30 June 1995.

Overall, the model is at a medium level of maturity. The system and workload representations are each at a somewhat limited level of maturity. These areas are not at a higher level of maturity because of missing workloads, system functionality, and computer and communication resource characterization. The Performance Statistics area is also at a somewhat limited level of maturity based on what evaluations and validations should have been performed for CDR and the lack of some performance statistics collection. The model structure is rated at a high level of maturity because the model is modularized and parameterized to a high degree and the BONeS tool is sufficiently flexible such that modifications can be made without unreasonably impacting the existing model components or without taking an excessive amount of time.

The categories shown in Exhibit 5-1 are further broken down into sub-categories, and these sub-categories are also given level of maturity values. The level of maturity values at the sub-category decomposition level determine the level of maturity values given in Exhibit 5-1, which in turn, determine the level of maturity values given in the executive summary for the Performance Model evaluation (i.e., the maturity level values are rolled-up from the bottom to the top). The level of maturity values at the sub-category decomposition level are presented in Appendix C. The remainder of this section discusses the evaluation results in the specific modeling analysis areas.

5.3.1.1 System Design Representation

The system representation in the Performance Model is evaluated for the inclusion of the subsystems of the design, the distributed sites, and the computer and communication resources of the design. The system representation is at a somewhat limited level of maturity because the delay characteristics of some of the hardware and software components of some of the subsystems are not represented. Exhibits C-2 through C-4 in Appendix C contain the details of the evaluation. A top level diagram of the Performance Model is shown in Exhibit 5-2. A mapping of the subsystem representations in the model is given on the next page.

Data Server: Represented by the Data Handler and Distribution modules.

Ingest: Represented by the Ingest module.

Interoperability: Represented by the inter- and intra-site network links (resources); Advertising and Subscription services of the Interoperability Subsystem are not represented.

Planning: Represented minimally in the Event Driven Scheduler module.

Data Processing: Represented by the Processing and User Processing modules.

Data Management: Not represented in the version of the model evaluated.

Client: There are no processing and I/O storage representations of the Client services; however, the Pull Generator module represents the Client transaction sources.

Model	Model Detail	Analysi Priority	Level Maturity	Analysis
1.	1.1	1	2	7/31/95
Representation	1.2 Resources	1	2	7/31/95
	1.3 Interfaces	1	2	7/31/95
2. Workloads	2.1	1	3	7/31/95
	2.2 Pull	1	1	7/31/95
	2.3 Reprocessing	1	0	7/31/95
	2.4 V0	2	2	7/31/95
	2.5 System	2	1	7/31/95
3. Performance	3.1 Resource Metrics	1	3	7/31/95
Statistics	3.2 Performance Requirements	1	0	7/31/95
	3.3 Performance Results	1	2	7/31/95
4. Model Structure	4.1 Modules	2	3	7/31/95
	4.2 Data Structures	2	3	7/31/95
	4.3 Parameters	2	3	7/31/95

Notes:

Analysis Priority:
1=highest, 3=lowest

Level of Maturity:
0=not addressed
1=limited maturity
2=somewhat limited maturity
3=fully mature
Blank=not yet analyzed by IV&V

Analysis Date:
Date that listed level of maturity was achieve
as reported in a TAR or TAM.

EXHIBIT 5-1: Performance Model Maturity Table

EXHIBIT 5-2: Top Level BONeS Model Diagram

The distributed DAAC sites are represented within the subsystem modules. Each subsystem is assumed to have its own set of computer resources. The sites identified in the model input are: ASF, EDC, GSFC, JPL, LaRC, MSFC, NSIDC, and ORNL.

The Performance Model represents computer and network resources for the subsystems at each DAAC. The inter-DAAC wide area network (WAN) and the intra-DAAC local area networks (LANs) are represented by communication links. The resource types for the DAAC computers consist of processors, working storage disks, archive stations (read/write heads), robots, and I/O channels. Two types of storage resources are represented: working storage RAID disks and archive tape devices with robots. A third tier of storage to data between the working storage disks and the archive tapes has been examined in a later version of the model.

Memory is assumed to be unconstrained and is not included in the model as a resource. Exhibit 5-3 shows the computer resource types represented in the major model modules. The network resources are included where transactions are routed between subsystems within a single site or between sites. The salient characteristics of the model's representation of the different resource types are given below.

Model Module	**Processors**	**Working Storage Disks**	**I/O Channels**	**Robots**	**Archive Stations**
Distribution		X	X		X*****
Data Handler		X	X	X	X
Ingest		X	X	X	X
Processing	X	X	X		
External		X	X		

***** The Distribution Subsystem includes storage devices for media distribution; these devices have the same representation in the model as Archive Stations

EXHIBIT 5-3: Computer Resources Types

Working Storage Disk

Data volume is the primary factor considered in the modeling of disks and is used to determine the amount of disk storage required. RAID-type disks are assumed for the model. Dataset mapping to disk drives is not represented.

Delay is not a factor in disk resource (i.e., drives and controllers) contention in the model representation. Disk delays are assumed to be overlapped in time with other service times (e.g., processors, network links, I/O channels, and archive tape drives) in the workload threads where data are written to or read from disks. However, disk delay can be included as a serial delay for production processing but is considered to be completely overlapped with processing time for the current version of the model. Disk delay is computed from file size, disk block size, access time, and disk transfer rate.

Archive Storage
Delays are modeled for robots and tape drives for archive storage. Robot delays are included for retrieving and returning tapes in the archive. Tape drive delays are included for mount, read/write, rewind, and dismount. Read/Write time is computed from file size and tape transfer rate. All delays are scheduled on a first-come, first-serve (FCFS) basis. There is no representation of data organization on tapes.

Network Links
Network link delay is represented by file size divided by link transfer rate with derating for protocol overhead. Protocol access methods, protocol processing, routers or communication devices are not modeled directly. WAN delays are modeled as point-to-point links using a time-slicing service discipline representing the packet level multiplexing of multiple sources on the link. LANs are modeled as dedicated links like HiPPI LANs . For workload threads where files are transferred from a source disk to a destination disk across a local or wide area network link, only the network link delay is represented. This corresponds to the assumption that the network link is the slowest server, and loading (and therefore queueing) is not significant on the other servers. There are cases in the design where the network link is not the slowest server (e.g., HiPPI).

Processors
Production processing is represented by FCFS processor delays for the science algorithms. Processors are assumed to be the SGI Power Challenge type with nominal capacity rated at 300 MFLOPS. Vendor MFLOPS ratings are derated by factor of four. Processing times are computed from the effective processor execution rate and the science algorithm floating point operation estimates from the AHWGP.

Processing of user Manipulate service requests are represented in the User Processing Module. Processing time is computed based on estimates of CPU service demands for simple, moderate, and complex transactions.

I/O Channels
Delays are not represented for I/O channels. They are assumed to be dedicated for the duration of delay of disk-to-disk transfer over a local or remote network link, and their service time is assumed to be overlapped by the other resource service times.

5.3.1.2 Workloads

The model's representation of workloads is at a medium level of maturity because several very significant workloads are missing. Exhibits C-5 through C-9 in Appendix C contain the details of the evaluation.

Three modules in the Performance Model (see Exhibit 5-2) generate the simulated transactions:

- Pull Generator
- Root/Ext File Generator
- V0 Generator

The Pull Generator samples from three distributions in generating pull transaction execution flows. These are described below:

- Diurnal Transaction Profile Distribution - determines service request frequency by time of day
- DAAC Distribution - determines the source DAAC where the service request occurs
- Service Type Distribution - determines which type of service is being requested for the selected DAAC (three types currently represented)

Currently, the estimates for these distributions are derived from the original HAIS user survey. The HAIS modeling team plans to update those estimates when new data are available from the EGSUS and AHWGC surveys.

The Pull Generator includes three service types: Ingest, Manipulate, and Exchange. There are numerous other service types, including six types of searches, that are not represented in the model (see Exhibit C-6 in Appendix C).

The push transactions are generated by the Root/Ext File Generator. The transactions are assumed to arrive periodically by either network transfer (i.e., from EDOS or PACOR II) or tape media. The amount of data for each network arrival is assumed to be a single data granule, which can be different sizes for different instruments. Loads are represented for TRMM and AM-1 instruments.

The following workloads are not represented currently:

- Reprocessing
- Subsetting/Subsetting (limited representation)
- non-EOS push data
- DAO/DAS
- IDS
- SCF QA
- MODIS Level 3 Products

5.3.1.3 Performance Statistics

The performance statistics area is at a somewhat limited level of maturity. Exhibits C-10 through C-12 in Appendix C contain the details of the evaluation. While the model has a high level of maturity for resource performance metrics, performance statistics for quantifying response times are not now collected. The statistical analysis methodology used to analyze the model data is sound; however, more results should be generated to support Release A SCDO CDR and more validation of results should be performed.

Resource utilization metrics are collected for all of the resources defined except I/O Channels. Total delay times are collected for production processing; however, no statistics collection probes are defined for end-to-end response time in the version of the model evaluated. Performance results from trade studies using the model are available to NASA, IV&V and the HAIS design team; however, some trade studies for CDR support will not be completed until after this TAR has been published.

The types of statistics collected by the Performance Model are summarized in Exhibit 5-4.

	Statistics				
Probe	**# Units**	**Queue Length**	**Data Volume**	**Throughput**	**Timing**
Processors	Timeline, Mean	Timeline			
Processes				# completed / day	Timeline, Mean, Min, Max, Std Dev
WS Disk			Timeline, Mean		
Archive Stations	Timeline, Mean	Timeline			
Robots	Timeline, Mean	Timeline			
WAN Link	Timeline, Mean				
LAN Link	Timeline, Mean				

EXHIBIT 5-4: Resource Metrics

The HAIS modeling team has used a variety of appropriate statistical analysis techniques to analyze the simulation data. Results for studies where multiple parameters have been varied over a set of model executions have been analyzed using Analysis of Variance (ANOVA) techniques. The SAS JMP statistical analysis package is used to perform these analyses. The ANOVA technique is the 2^n factorial design where n factors are studied at high and low values. The model is executed for all combinations of factor levels and the data (i.e., model results) are analyzed using the ANOVA technique. This technique is especially appropriate for separating the effects of the different factors on the response variable (i.e., the statistics collected by the model) and any interaction between the factors on the response variable.

All other results are simple statistical computations (e.g., mean, minimum, maximum, standard deviation) computed within the BONeS probe functions. The statistical analysis techniques implemented in the BONeS probes are appropriate for the cases where they are used. Confidence intervals for mean values have not been computed on the results reviewed to date.

Exhibits 5-5 and 5-6 give lists of results received by the end of June and studies planned for Release A CDR and Release B IDR, respectively. Note that the missing workloads previously listed correspond to Release A SCDO requirements but are not being evaluated using the Performance Model until the Release B IDR timeframe. Results from the studies planned for Release A that are not already available will not be available to us until August at the earliest. The Release B studies will be performed in the September-October timeframe.

Study Objective	**Push Load**	**Pull Load**	**V0 Load**
Third Tier Storage	Epoch k: 1) CERES only; 2) all instruments	Yes	Yes
2 Shifts/Day	Epoch e	No	No
MODIS Level 2	MODIS	No	No
Archive Technology	Epoch k	Yes	Yes
Platform Hardware I	Epoch e	Yes	Yes
Platform Hardware II	Epoch k	Yes	Yes
Hardware Failure/Recovery	Epoch e	Yes	Yes
Release A Sizing	Epoch e	Yes	Yes
Release B Sizing	Epoch k	Yes	Yes
Production Topologies	Epoch f, LaRC	No	No
CERES Staging	Epoch k, CERES	No	No

EXHIBIT 5-5: Model Results Completed

Release	Study Objective
A	Non-EOS Data: Landsat-7, ADEOS II, Meteor, ACRIMSAT, ALTRADAR, COLOR Updated AHWGP Data (including MODIS Level 3 Data) DAO IDS Survey Data (Interim) EGSUS Survey Data (Interim) Baseline Combined Runs End-to-End Response Times Hardware Failure/Recovery LAN Utilization WAN Utilization
B	Updated AHWGP Data SCF QA & Science Data DAO/DAS IDS Survey Data (Final) EGSUS Survey Data (Final) Test/EDF User Supplied Methods Platform Resources Hardware Failure/Recovery Guaranteed Service Data Manager User Characteristics Sensitivity LAN Utilization WAN Utilization Reprocessing

EXHIBIT 5-6: Model Studies Planned

HAIS has performed limited validation of model results using analytical techniques. Only utilization metrics have been computed for the average and peak number of resource units in use as shown in Exhibit 5-7. The statistics in the exhibit are for ECS totals. The results show good agreement except for the disks. It should be noted that these utilization-type metrics can be estimated to within a few percent error using analytical techniques. In the cases where agreement is not good, HAIS cites as the reason the fact that the analytical technique does not take into account the additional time that the data is kept on disk before it is deleted (i.e., beyond the time the data are in active use by a process).

	Avg # CPUs	Avg Proc Disk (GB)	Peak Proc Disk (GB)	Avg DH Disk (GB)	Peak DH Disk (GB)	Avg #Robots	Avg # Archives
Static Model	113.61	499.14	695.32	52.39	251.07	14.25	41.15
Dynamic Model	113.67	631.72	804.76	30.25	246.92	14.03	39.96
Absolute Error	-0.06	-132.58	-109.44	22.15	4.15	0.22	1.19
Relative Error	-0.05%	-23.45%	-14.59%	53.59%	1.67%	1.53%	2.94%

<u>Abbreviations</u>
DH - Data Handler (corresponds to Data Server Subsystem   Avg - Average
Proc - Data Processing Subsystem   GB - Gigabytes

EXHIBIT 5-7: Validation Results

HAIS plans to perform a calibration of the simulation model during the Release B study time frame (i.e., September - October). The model will be calibrated to measurement results obtained from EDF benchmarks.

At this point we have executed the model for a limited number of cases and have been able to reproduce some of the results.

5.3.1.4 Model Structure

In general, the structure of the ECS Performance Model is at a high level of maturity. Subsystems are represented in modules. Resources within subsystem modules and the distributed DAAC sites are represented parametrically. The characteristics of transactions are implemented in terms of data structures that are accessed by the logic contained in the module definitions. These features will aid in minimizing modification effort. Many workload and service demand parameters for the modules are read from files or provided via screen input. Consequently, many changes can be made simply by data input. However, wholesale changes to the ECS architecture (e.g., processing hub and distribution centers) will require significant effort in modifying the model (although many of the modules could be re-used). Exhibits C-13 through C-15 in Appendix C contain the details of the evaluation.

Exhibit 5-2 shows the actual BONeS block diagram for the highest level of the Performance Model. The BONeS system modules at the top of the diagram represent the major SCDO subsystems; the push, pull, and V0 workloads, and other necessary modeling functions. Each module in the diagram has many more block diagrams implicit in its definition. There are thousands of modules in the model. The diagram shows a probe module (at the top of the diagram) that implements standard BONeS-defined and user-defined statistics collection functions. The diagram also shows model parameters (represented by the letter ‘P’), system resources (represented by the letter ‘R’), and model variables that are shared across multiple modules, called memory (represented by the letter ‘M’). Model parameters, resources, and memory can be defined for each module level. Parameter values are provided from file input or screen input. The data structure definitions are not shown here because they are quite lengthy. They define the characteristics of the files and processes for the workload transactions.

5.3.2 Identified Problems

The problems identified for the ECS Performance Model are divided into two categories: 1) problems where HAIS has defined a plan to address the problems, and 2) problems where HAIS has not defined a plan. The problems for these two categories are discussed below.

In the results discussion above several deficiencies are identified for missing workloads, additional model validation, and additional performance evaluations. HAIS expects to address the following deficient areas prior to the Release B IDR:

- End-To-End Response Time - to compare with response time requirements
- LAN/WAN utilization - to size networks
- Test/EDF model - to size EDF components and calibrate the model
- Reprocessing Workload - updated AHWGP data to enhance push workloads
- Subsetting/Subsampling - updated AHWGP data to enhance push workloads
- SCF QA data - updated AHWGP data to enhance push workloads
- Non-EOS push data - new data to complete push workloads
- DAO/DAS data - new data to complete push workloads
- Hardware failure/recovery - to examine failure scenarios and recovery times
- IDS Survey - new AHWGC data to enhance pull workloads
- EGSUS user survey - new data to enhance pull workloads
- User supplied methods- new data to enhance pull workloads
- Data Management Subsystem - model enhancement of pull workloads

The above list of enhancements to the model and the additional evaluations will feed into the hardware sizing exercise. The approach apparently being used by the HAIS design team to compensate for workloads not included in the model is to extrapolate from the model resource utilization results. For example the number of processors estimated by the model for the Data Processing Subsystem is probably tripled since Reprocessing is estimated to be twice the load of direct Processing which is represented in the model. This is a reasonable technique to determine the number of devices since utilization is a linear function of load; however, delay and throughput metrics are not linear functions of load and cannot be assessed by the same analytic technique. One risk inherent in this approach is that the number of resources (e.g., processors) determined based on a utilization metric may not be sufficient to meet the system delay requirements. The importance of this delay evaluation is underscored because a basic assumption of the model is that the resources are at 100 percent utilization during peak periods. Since delay increases exponentially as utilization approaches 100 percent , it is very important to evaluate the system delays and evaluate resource capacity margins where **all** of the workloads and service demands are included (e.g., a hardware failure simulation showed that the system will not be able to catch-up with production processing after an eight hour outage). The risk in not doing so is that system resources and cost will be underestimated.

There are other significant performance issues (some more significant than others) that need to be evaluated. These issues are summarized in Exhibit 5-8. There is currently no HAIS plan that we know of to evaluate the following issues.

Memory is not represented in the Performance Model for several reasons. One reason is that the detailed data required to model memory management of processes and data are not available. Another reason is that it is not practical to include memory in a system level model because of the expected model execution time. However, given the process execution times and the quantities of data involved in production processing, process or data swapping between memory and disk could degrade performance significantly in the Data Processing Subsystem. Memory contention in the Data Server Subsystem could also be a problem because of concurrent file accesses and transfers.

Closely coupled with the memory sizing issue is the issue of process scheduling in the Data Processing Subsystem. The model assumes that process execution is scheduled on a FCFS basis. While this service discipline may work better than a time-sliced discipline to limited swapping of processes and data, it will result in longer queueing times for processes with shorter service time requirements.

There are several areas in the model that are deficient in representing delays as pointed out in the results discussion above. The potential impact of not examining the delays in the model is that hardware resource sizing may be underestimated. The areas where more fidelity is warranted in the model are summarized in Exhibit 5-8**.**

Problem Area	Problem Characteristics
Memory Sizing Analysis	Process & data memory requirements missing Process scheduling discipline will affect delay & memory requirements
More fidelity in modeling delays	Assumption that Data Processing Subsystem I/O is 100% overlapped with processing Assumption that network link is slowest server in disk-disk thread Representation of storage I/O missing for User Processing Contribution of delay for disk drives and controllers for working storage Contribution of delay for CPUs (non-Data Processing Subsystem) Overhead like protocol processing, DBMS, & file mgt S/W

EXHIBIT 5-8: Problem Areas and Characterizations

5.3.3 Potential Issues

A potential issue in the performance evaluation of the ECS design is that there are no estimates of non-science user workloads. The AHWGC and EGSUS surveys are targeted towards science users. The frequency and characteristics of non-science user access is unknown and, as a consequence, is not being taken into consideration in the design. The risk is that the system will not have the capacity to service this set of users.

Short of a separate characterization and evaluation of this workload, this potential issue can be examined by increasing the arrival rate for the science user pull workload.

5.4 Conclusions

5.4.1 Technical Integrity

The ECS Performance Model is reaching a level of maturity where its results can be trusted for some purposes. The caveat in using the results is that the reader must understand the inherent deficiencies. The technical integrity conclusions for the Performance Model evaluation are discussed below for the evaluation metric categories presented in Section 5.1.2.

5.4.1.1 Engineering Quality Metric

Because the delay characteristics of some of the hardware and software components of some of the subsystems are not represented, several performance issues cannot be examined. The outcome of the analyses of these performance issues may impact performance requirements compliance and cost. However, the level of detail in the model is reasonable to assess the primary drivers that impact cost at the detailed design level once enhancements are made and all workloads are considered in the performance analysis.

The model assesses processing requirements only for the Data Processing Subsystem. The processing demand for the Ingest data input (e.g., for interrupt handling and working storage input/output processing) is bound to be significant because of the high data rate. The processing delays for Ingest and the other subsystems are serial delays and contribute to processor loading. They need to be included to adequately size the processors for costing purposes and to assess the performance requirements compliance.

Adequate estimation of system overhead performance parameters has not been made at this point. Network protocols like FTP, TCP/IP, remote procedure calls, and data input/output functions (e.g., database and file management software), will cause performance degradation.

5.4.1.2 Testability Metric

The model is reaching a level of maturity where its results can be validated. The early validation results reported here represent a first step. The planned model calibration exercise with EDF measurement data during the Release B analysis timeframe will provide important data for refining parametric input to the model as well as for validating output produced by the model.

5.4.1.3 Traceability Metric

The model is still not at a level of maturity such that performance requirements compliance can be assessed. This deficient area can be remedied with minimal effort.

5.4.1.4 Flexibility Metric

The model is very mature in this area, and the HAIS modeling team should be able to respond to requests from the different groups who use model results. Since the SDPS PDR, many modifications have been made within the same basic model structure to address different design issues. The limitation in performing trade studies with the model has typically been due to the lack of data from outside groups instead of the modeling team not being able to adapt the model to perform the studies.

5.4.2 User Satisfaction

The design team is now interacting more with the modeling team. It is clear that the results are useful to the design team. They have been requesting trade studies and are apparently making decisions based on model results. The design team could be using the model to even greater advantage if they would provide the modeling team with more detailed design data and evaluation criteria to use in the model. (e.g., response time requirements have not been evaluated, in part, because the design team has not provided the detailed information necessary to perform this analysis).

Instrument teams have also been very actively interacting with the modeling team. They have revised their concepts of operations in some cases based on modeling feedback. The most notable cases are:

- CERES data staging study
- MODIS production of Level 2 and Level 3 products studies

Each time a new set of data is provided to the modeling team, a significant effort is required by the modeling team to convert the data into model input because of errors in the data. The instrument teams and the HAIS science team should perform better QA on the data provided to the modeling team to avoid schedule impacts in completing the modeling studies.

The third potential user of the model, the ESDIS Project Office, has had no direct visibility into the model but has received all of the results referenced in Exhibit 5-5. They have however, received results of the IV&V modeling assessment conducted prior to PDR [3].

5.4.3 Trends and Projections

The Performance Model is at a somewhat limited level of maturity, up from a limited level of maturity at PDR. We expect it to be at a higher level of maturity by Release B IDR. The model contains more workload detail but has the same basic representation of the system design. There are differences in production processing and user requests, and there are plans to add more subsystem functionality and workloads. Useful results are being produced and used by design and instrument teams to good advantage. However, the model still has a lot of room for improvement in maturity, and more studies should be performed.

5.5 Recommendations

5.5.1 Areas Requiring Further Analysis

We recommend that additional Release A studies not yet available and Release B studies be analyzed. This will entail not only analyzing the actual results but also the model modifications made to produce the results.

Additional validation of results using the model is another area requiring further analysis. At this point we have performed limited validation of the results.

5.5.2 Solutions To Important Problems

Two important problems identified above are the lack of memory sizing and the need for more fidelity in assessing delays. Approaches are outlined below for these problem areas.

Two approaches are outlined for performing a memory sizing analysis. One approach is based on using a variant of the system Performance Model; the other approach is based on empirical measurement and statistical analysis.

The steps of the modeling approach are as follows:

- Construct a BONeS model from a subset of the system model components (e.g., Data Processing Subsystem)
- Collect data from the system model execution to characterize the arrival profile of requests to the subsystem model by process id for different peak periods (e.g., weekly and monthly production processing)
- Convert arrival frequency data into cumulative distribution function (CDF) for all processes, one CDF for each peak period
- Characterize resource service demands including memory requirements by process id and convert into CDF for all processes
- Add memory resource arguments and logic to the subsystem model
- Add workload generator that samples from arrival frequency and process memory requirements CDF's
- Change the processor resources in the subsystem model to Server Resources (from Quantity Shared Resources) and use Processor Sharing service discipline as well as FCFS
- Design different experiments with different amounts of memory and different processor service disciplines
- Execute the model and collect memory resource statistics
- Analyze the statistical data and determine the appropriate amount of memory

The steps of the empirical memory sizing approach are as follows:

- Execute the system model and collect the transaction profile by process id for the subsystem being evaluated
- Develop (or procure) a synthetic workload generator
- Obtain real or representative science algorithms and install them on the EDF Unix system
- Design several experiments with different amounts of memory and different operating system settings for process scheduling
- Using the synthetic workload generator, execute the model script on the EDF system for the different memory and process scheduling alternatives
- Collect memory usage statistics using a Unix monitor
- Analyze the statistical data and determine the appropriate amount of memory

The deficient areas of delay representation in the model can be addressed in a straight-forward manner by adding resource service demands. The following changes would be necessary:

- Add processor resources for the Ingest and Data Server subsystems
- Add protocol, database management, and file management software processing time estimates throughout
- Add disk controller resources for all subsystems
- Add disk delays where not already characterized for push and pull transactions

- Modify workload flow logic to include the new resources
- Add appropriate statistics probes to collect response time data

5.5.3 Risk Management

Accurately estimating cost is a critical aspect of the ECS program. A significant part of the inputs to the Cost Model come from the Performance Model results. Therefore, the Performance Model's results should be as accurate as possible. To minimize the risk of inappropriate sizing, and thus inaccurate cost estimation of the system, we recommend that the problem areas identified be examined as soon as possible.

We view modeling as a risk reduction methodology. We recommend that the evaluation of the modeling activities continue to insure that all performance issues get addressed.

This page intentionally left blank.

6. COST MODEL

The Cost Model is intended to estimate the resources required to develop and operate ECS architecture alternatives (as partially derived from the Performance Model) within schedule constraints. The Cost Model is currently implemented as a collection of three types of stand-alone cost estimation methods: Custom Software, Operations and Maintenance, and Commercial-Off-The-Shelf (COTS) Hardware and Software. The findings of an independent assessment of the cost modeling activities in each of these three areas including analysis tasks performed, constraints affecting the analysis, results, conclusions, recommendations, and risk management are presented in this section.

6.1 Analysis Tasks Performed

In the first IV&V Modeling Assessment TAR [3], a thorough evaluation of the individual cost modeling areas was performed. That analysis examined the Custom Software, Operations and Maintenance, and COTS Hardware and Software cost estimation methods employed by the ECS contractor (i.e., HAIS). These three estimation methods correspond to the cost breakdown being used by HAIS and encompass the vast majority of the costs associated with developing the system. Rather than reiterate the findings of the previous report, this analysis identifies the salient portions of that report and presents new findings. The items analyzed for each modeling area are described below.

6.1.1 COTS HW and SW Estimation

Three models have been utilized by HAIS in the estimation of COTS hardware and software costs. They are:

- The COTS cost estimation model
- The COTS procurement model (a.k.a. the Bill-of-Materials Cost Model)
- The distribution cost sensitivities model

Our analysis focuses primarily on the COTS procurement model and the COTS estimation model. The distribution cost sensitivities model was not evaluated in this report. The COTS procurement model identifies the number of components, the cost per component, the overall cost trends, and the reasonableness of those cost trends. The COTS estimation model is a tool designed to perform trade-off analysis with relative speed and ease, while accurately reflecting various cost scenarios.

Included in our analysis is the evaluation of HAIS' Interactive Cost Model (ICM) [8,9]. The intent of this model is to provide the EOS scientist a tool to construct "What if" scenarios for different processing requirements. The output from the model is a "Cost by Fiscal Year" exhibit showing the cost from 1994 through October 2003 for processing, RAID disk, archive, maintenance, and a total cost by fiscal year. The COTS estimation model and COTS procurement model are described in Appendix D.

6.1.2 Custom SW Estimation

Like our previous analysis, our analysis examines the methods used by HAIS to estimate the creation of custom software. Their method involves an initial estimation of the software size from which schedule and cost estimates are derived. A complete description of the method used to estimate size, effort, and schedule is found in Appendix D.

We examined custom software estimation methods used by commercial software developers in order to validate the methodology employed by HAIS. Some of the specific tasks performed include:

- Search for an established relationship between source lines of code (SLOC) and (OOD) entities
- Empirically determine the relationship between SLOC and OOD entities such as objects and methods in existing C++ software
- Determine alternative methods/tools capable of validating the SLOC to OOD HAIS is proposing

6.1.3 Operations and Maintenance Cost Estimation

Operational costs are driven primarily by personnel costs. Therefore, verifying HAIS' staffing profile for completeness and accuracy is key to ensuring their operations and maintenance cost estimation methods are sound. We attempted to verify the staffing profile acquired during our previous analysis. We also requested an updated staffing profile, along with other relevant data such as estimated cost per position for the CDR time frame. We intended to incorporate any information regarding the use of subcontractors and temporary personnel in filling some clerical positions. Because of the non-availability of any models or updated costing profiles, our analysis was limited to whatever new information we could glean through conversation with the HAIS representative responsible for operational cost estimation.

6.1.4 Model Interfaces and Design

Many of the cost estimates for ECS are a direct result of modeling activities. Our initial plan was to receive the access to the models and run independent tests to determine the flexibility of the design, the ease of use, and how the various models interface. This did not occur as planned because, with the exception of the ICM, IV&V was not provided access to the models. We ran independent tests on the ICM and evaluated the design and interface. The testing consisted of altering the data within the 15 input areas and examining the results. Interaction with the model developer helped us to gained a thorough understanding of the model.

6.2 Constraints Affecting the Analysis

Our initial plan was to test the cost estimation models within each area, and verify the output from each. With the exception of the ICM, we were denied access to the models. Therefore, our means of information collection was limited to personal interviews and electronic correspondence with the model developers at HAIS. Overall, the non-availability of models and cost information severely limited the degree of analysis that could be performed. Although the HAIS staff are knowledgeable and experienced, and their comments and described approaches seemed reasonable, it is strictly impossible to verify the accuracy of those comments without access to the models or the underlying cost information. The only other approach is to perform independent estimates, which is altogether a separate task. Also, without concrete information, the analysis is limited to examining the validity of the stated approaches; no analysis of cost accuracy is possible. Generally, little progress in obtaining the necessary information was achieved since our first report. The specific constraints encountered within each modeling area are described below.

6.2.1 COTS HW and SW Estimation

Within the COTS hardware and software modeling area, the primary constraint was the non-availability of models. With the exception of the ICM, the IV&V team was not given access to any of the COTS models. The lack of access to the models and supporting cost information made a complete and thorough analysis impossible.

6.2.2 Custom SW Estimation

Constraints encountered in the custom software estimation area are basically the same as those encountered during our previous analysis. They are as follows:

- The very nature of this type of estimation is a constraint. Traditional LOC estimating techniques can yield misleading or erroneous size, cost and schedule results when they are used to estimate OOD software.
- The detailed design of the software has not been completely solidified.
- Producing counts of software entities is a highly subjective process, and validation of that process is more difficult at this point than it would be in later stages of the lifecycle.

6.2.3 Operations and Maintenance Cost Estimation

Our previous analysis revealed that personnel estimates were performed manually and based upon experience rather than upon established metrics. We could not find evidence that suggested this method of estimating has been appreciably altered. Although this method may be typical for this type of cost estimation, the result is a method that is labor intensive and error prone.

6.3 Results

It should be noted that the results reported here assume that the information provided in the interviews and correspondence with the ECS contractor is accurate. Very little hard or soft copy data was made available to cross check the information gleaned from the interview process.

6.3.1 Discussion of Results

The results of the Cost Model evaluation are discussed below by the four evaluation areas: COTS hardware and software, custom software, operations and maintenance, and model interfaces and design. The evaluation is summarized in Exhibit 6-1. As shown in the exhibit, each evaluation area is broken into multiple categories; each category is given an analysis priority, level of maturity, and analysis date.

Cost Model: Top Level				
Model	**Model**	**Analysis**	**Level of**	**Analysis**
1. COTS Hardware and Software	1.1 COTS Cost Estimation Model	1	2	31-Jul-95
	1.2 COTS Procurement	1	2	31-Jul-95
	1.3 Distribution Cost Sensitivities Model			
2. Custom Software	2.1 Size Estimates	1	2	31-Jul-95
	2.2 Level of Effort/Schedule Estimates	2	2	31-Jul-95
	2.3 Cost Estimates	2	2	31-Jul-95
3. Operations and Maintenance	3.1 Staffing Estimates for Each Entity	1	2	31-Jul-95
	3.2 Staffing Cost Factors	2	2	31-Jul-95
4. Model Interfaces and Design	4.1 User Interfaces	1	1	31-Jul-95
	4.2 Interfaces Between CM	2	1	31-Jul-95
	4.3 Interfaces Between 3 COTS	1	1	31-Jul-95
	4.4 Interfaces With Performance	1	1	31-Jul-95

EXHIBIT 6-1: Cost Model Maturity Table

6.3.1.1 COTS HW and SW Estimation

The cost profiles developed for CDR will be created using the same set of models that were used for the previous costing exercises. Our findings for each of these models are described below.

COTS Cost Estimation Model - This model estimates costs for COTS hardware and software, required maintenance, and associated operations costs through the end of the contract. It uses a 25% efficiency factor in computing the number of components required. This is a contract requirement. When computing hardware and software costs directly (rather than as a ratio of current capacity estimates to previous capacity estimates multiplied by previous cost estimates), it uses standard multipliers to account for costs associated with hardware maintenance and operation of the processors. The multiplier for hardware maintenance is 9.5%. Our analysis revealed that the 25% factor is being applied correctly and the 9.5% maintenance and operations factors are reasonable.

COTS Procurement Model - The COTS Procurement Cost Model, also known as the Bill-of-Materials Cost Model, is a means of estimating COTS hardware and software costs given a bill-of-materials. The output of the model is the total COTS hardware and software cost across the life of the project for the input bill-of materials. This model also has the capability to compute maintenance costs as a function of time. Costs in the COTS Procurement Cost Model are updated whenever repricing activities (such as Change Order #1 updates) occur. This has occurred at least four times since the start of the contract (August 1993, April 1994, for PDR, and most recently for CDR).

Model Implementation - The COTS cost estimation model has been implemented as several Excel spreadsheets. The COTS procurement model is implemented as a series of dBASE files (database files and program files). Running the COTS cost estimation model is a manually intensive process. It is incumbent on the operator to account for all functions, sites, changes, etc., and insure that all relevant costs are included. Moreover, the spreadsheets used to implement the model are not linked. The operator must manually cut-and-paste outputs from one sheet into another.

Our analysis of the Interactive Cost Model (ICM) revealed that it was essentially an abbreviated version of the COTS estimation model, with some modification to account for pull requirements. From a strictly mechanical standpoint, the ICM developers followed a reasonable and systematic approach. Details of our analysis of the ICM are documented in two Technical Analysis Memoranda (TAMs) [8,9].

We consider the COTS hardware and software estimation models to be of somewhat limited maturity for the following reasons:

- The models appear to be manually intensive
- The Excel spreadsheets of the COTS estimation models are not linked
- The models contain a number of inputs that must be verified each time they produce an estimate
- We did not receive access to the models, to verify their functionality

6.3.1.2 Custom SW Estimation

The method used by HAIS to estimate software size is as follows:

1. Count objects/classes, as contained in the System Design Specification
2. Characterize each object/class as simple, average, or complex
3. Estimate the numbers of operations per object/class
4. Multiply the number of operations in each object/class by fixed numbers of SLOC based on the complexity characterization:
   simple 100 SLOC / operation
   average 150 SLOC / operation
   complex 200 SLOC / operation
5. Sum the number of SLOC across all objects/classes
6. Add a 10% margin of safety

Schedule and cost estimates are created from the output of the size estimate. A complete description of HAIS' custom software estimating process is described in Appendix D.

Software estimation is a very imprecise science hence it is prudent to perform the software estimate carefully and consider multiple approaches. It appears that HAIS has yet to verify their software cost and schedule estimates using another software estimating technique. Therefore, it is difficult to determine whether HAIS' custom software estimation method is producing reasonable results. As reported in our previous analysis [3], Reifer Consultants, Inc., a group of software development experts, states that software estimation can be considered complete when and only when:

- two estimates have been done using different techniques
- the two estimates have been compared and any major differences explained
- the final estimate has been verified
- the baseline estimate is established and documented

Methods used to estimate schedule and cost aspects were well documented and defined. As stated above however, the schedule and cost estimates are dependent on the size estimate. Since the method used to estimate size is suspect, we consider the custom software estimation model to be of somewhat limited maturity.

6.3.1.3 Operations and Maintenance Cost Estimation

Operations and maintenance cost estimation includes costs in the following categories:

- Operations staff
- Maintenance staff
- Integrated Logistics Support
- Sustaining engineering
- Management
- Training

The general process of estimating staffing levels is: 1) identify staffing requirements for each system function, 2) determine the size of each operational entity (from modeling results) and associated workload, 3) determine size of code, 4) estimate staff, and 5) allocate staff. Although there are areas in which the number of staff is estimated according to a mathematical formula, the process is not quantitative in general. Staffing levels, allocation of staff to locations, and allocation of responsibilities to staff are primarily determined based on experience and judgment. In some instances quantitative measures are used, such as the number of lines of code maintainable by one person, time to handle a single media (including mounting, dismounting,) and the cost for hardware maintenance. However, according to HAIS, a salary of $100K per man year (regardless of position) is assumed for costing purposes. Appendix D contains the results of a re-verified personnel costing exercise.

Our conclusion regarding the quality of the operational cost estimation area is that the approach is reasonable. The cost estimation process is thorough; all of the elements and major staff functions have been included. There is concern regarding the lack of automation of this modeling area, leading to overly subjective cost estimates and an inability to support trade analysis. Therefore, we consider this area to be of somewhat limited maturity.

6.3.1.4 Model Interfaces and Design

Without access to the models, it is difficult to verify their design or interface functions. A great deal of information was learned regarding the models, their design and functions from our interviews and correspondence with HAIS personal. However, without access to the models to verify the assertions of the HAIS personnel, we consider the model interfaces and design area to be of limited maturity.

6.3.2 Identified Problems

In general, few aspects of the modeling activities were determined to be seriously flawed. The lack of information prevents the IV&V team from reaching firm conclusions in a number of areas. Discussion of specific problem areas follow.

6.3.2.1 COTS HW and SW Estimation

In the COTS hardware and software estimation area, three problems were identified regarding the COTS Cost Estimation Model. These are the same problems identified in the first report [3] and are summarized as follows:

- Operating the models is a manually intensive process
- Using old cost data to determine future costs leads to inflated estimates
- The parameters used in the COTS cost estimation model are too conservative

An example of these conservative parameters is found in the compression assumption. The model assumes an archive compression ratio of 1.5. Two independent studies have found that the minimum compression ratio achieved using lossless compression of satellite data is about 1.7. However, compression ratios as high as 5 were demonstrated, with the average well in excess of 2.

6.3.2.2 Custom SW Estimation

Reifer Consultants, Inc. strongly recommend performing two estimates. To our knowledge, this has not been done.

6.3.2.3 Operations and Maintenance Cost Estimation

With regard to operations and maintenance costs, a key issue is the level of automation. Currently, this is the primary deficiency in the estimation of operational costs. The process is manually intensive, and as a result is only conducted once in each development phase. A separate method which estimates approximate operations costs in a semi-automated fashion is needed to facilitate trade analysis.

6.3.2.4 Model Interfaces and Design

The majority of costs associated with the ECS are generated via some type of model. Therefore, it seems logical to conduct a complete and thorough analysis of the models, and substantiate their design, parameters, interfaces, and output. Otherwise, the potential exists for reliance on the models when in fact their interfaces may not operate as intended and their basic design may fail to meet the requirements needed to produce accurate estimates. The non-availability of the models is a major problem impacting our analysis.

6.3.3 Potential Issues

In general, the potential issues for the cost modeling area are as follows:

- There is no lifecycle cost model
- The parameters being used in the models are generally too conservative
- The software estimation process is untested in the industry
- The operational estimates for personnel are manually intensive
- The model interfaces and design remain untested by the IV&V team

Each one of the items identified could cause the estimates HAIS is generating to overstate or understate the estimated costs of creating and operating the ECS.

6.4 Conclusions

Overall, the parameters being used in the various models appear to be too conservative. Within COTS hardware and software estimation, the prime example is the percentage decrease in cost per unit of performance per year. Within custom software estimation it is the SLOC multipliers per operation. Within operations cost estimation it is the cost per man-year.

6.4.1 Technical Integrity

It appears the engineering quality with respect to the models is good. Each cost modeling area was individually evaluated. The lack of access to the models and assumptions used to create the cost estimate precluded us from a comprehensive analysis.

6.4.1.1 COTS HW and SW Estimation

This modeling area would be moderately testable if the models were provided. Our analysis revealed that the spreadsheets that combine to create the COTS estimation model are not tied together in an automated fashion. Without direct access to the models a complete analysis is not possible.

6.4.1.2 Custom SW Estimation

The custom software estimation approach is moderately testable given the methods and the approach HAIS is proposing. However, without a validation of that estimate, the results are questionable. The HAIS technique for estimating software size seems reasonable, however it remains unverified and is not known to be calibrated. The method for estimating schedule and cost on the other hand, are more defined.

6.4.1.3 Operations and Maintenance Cost Estimation

The testability of this modeling area is limited. There are two main reasons for this. First, the methods that are being used by HAIS to determine costs in this area seem primarily driven by experience and judgment, and less so on established metrics. Hence, the primary means to validate the modeling area is through an independent analysis and comparison of results. This leads to the second reason for limited testability; the lack of detail describing the phasing of automation that is required to perform an independent analysis of operations costs.

6.4.1.4 Model Interfaces and Design

The purpose of this category was to address the models based on their operational effectiveness. We were unable to examine the models directly, other than the ICM. Through the interview process we have been able to determine that the COTS models were carefully designed, and seem to operate as intended.

6.4.2 User Satisfaction

6.4.2.1 COTS HW and SW Estimation

The potential users of the cost models are the engineering staff and the instrument teams (to support their trade-off analysis). The COTS cost estimation model satisfies the basic needs of the engineering staff, but limits the scope of their work due to lack of automation. The model implementation does provide maximum flexibility. Flexibility is important, because of the varying analyses required and because the ground rules are in a constant state of flux. This approach is dependent on the knowledge, analytical skills, and thoroughness of the operator. This has two effects. First, it makes the model essentially unusable by others. Second, it makes the execution of the model very time consuming. Further, the model currently contains embedded rate information, making its release to outside entities problematic. The COTS cost estimation model need not contain this type of information in order to support trade analyses. The accuracy of the estimates produced by the COTS procurement cost model, which is its most important requirement, could not be evaluated in this study due to a lack of available information.

6.4.2.2 Custom SW Estimation

The methods being used for custom software estimation certainly satisfies the needs of the engineering team. However, this model is inaccessible to all groups outside the engineering team due to the degree of detailed knowledge required. A simpler method is needed to support trade analysis and lifecycle cost estimation.

6.4.2.3 Operations and Maintenance Cost Estimation

As with the COTS hardware and software estimation techniques, the methods being used for operations and maintenance cost estimation meet the basic need of the engineering staff, but limit the types of analyses than can be performed. This model is inappropriate for general use due to the degree of detailed knowledge required.

6.4.2.4 Model Interfaces and Design

We are unable to verify the design and interfaces of the models due to the lack of information regarding the interfaces and the unavailability of the models.

6.4.3 Trends and Projections

Based on the limited information received since our previous analysis, there is no indication that the degree of automation, scope or methodology used within the models is changing.

6.5 Recommendations

6.5.1 Areas Requiring Further Analysis

Delivery of the cost models to the IV&V team along with the supporting cost information would make a more meaningful analysis possible. We recommend performing an independent custom software estimate using another estimating tool. Some specific tools are suggested in Appendix D. Once the size is solidified, we strongly suggest recalculating the schedule and cost estimates based on the validated size estimate. The cost elements used in this model should be updated where necessary. The staffing estimates should be refined based upon the latest staffing information and should use more realistic cost parameters, rather than the 100K which has been used throughout this costing exercise.

6.5.2 Solutions to Important Problems

There are three overall problems which can and should be addressed. First, automate the models. Automation would allow estimates to be produced more quickly and enable more “What if” and trade-off analysis. Second, make the cost models available to the IV&V team. Third, make the cost data available to the IV&V team. Release of the models and supporting cost data to the IV&V team would allow the government to receive an unbiased assessment of the costs HAIS has estimated to implement and operate ECS.

6.5.3 Risk Management

With a contract the size of ECS it is important to mitigate the risk through complete and thorough cost modeling effort and the use of current cost information. Therefore, contractors estimates and all supporting data should be carefully analyzed to ensure that all estimation techniques and the cost data used are reasonable. The primary risk in the COTS modeling area is that a fundamental design concept may fail, resulting in dramatic changes in the types of hardware required to implement the system. Mitigation of this type of risk involves more prototyping and modeling. The risk surrounding custom software development can be reduced through prototyping and validation of primary estimates via a second estimate. This reduces the risk of underestimating the cost and effort required to deliver a completely integrated and functional custom software program.

This page intentionally left blank

APPENDIX A: USER MODEL ANALYSIS DETAIL

A.1 User Demography Tables

The information used to create exhibits 3-2 , 3-4 , and 3-5 is displayed in exhibits A-1, A-2, and A-3.

Respondents in EGSUS Database (288 records)

Area	User Discipline Selection											
	Primary		Secondary 1		Secondary 2		Secondary 3		Secondary 4		Total	
atmos	75	34.09%	66	44.59%	0	0.00%	0	0.00%	0	0.00%	141	33.65%
land	54	24.55%	26	17.57%	14	35.90%	4	33.33%	0	0.00%	98	23.39%
ocean	61	27.73%	34	22.97%	11	28.21%	0	0.00%	0	0.00%	106	25.30%
cryo	12	5.45%	6	4.05%	1	2.56%	6	50.00%	1	100.00%	25	5.97%
public	4	1.82%	16	10.81%	13	33.33%	2	16.67%	0	0.00%	35	8.35%
other	14	6.36%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	14	3.34%
none	68	—	—	—	—	—	—	—	—	—	68	—
totals	288	100%	148	100%	39	100%	12	100%	1	100%	487	100%

EXHIBIT A-1: User Discipline Analysis: EGSUS Only (See exhibit 3-2)

Respondents Common to EGSUS and NASA HQ Databases (127 Records)

Area	User Discipline Selection									
	Primary		Secondary 1		Secondary 2		Secondary 3		Total	
atmos	35	35.00%	31	45.59%	0	0.00%	0	0.00%	66	34.74%
land	22	22.00%	15	22.06%	7	41.18%	2	40.00%	46	24.21%
ocean	25	25.00%	15	22.06%	4	23.53%	0	0.00%	44	23.16%
cryo	8	8.00%	3	4.41%	0	0.00%	3	60.00%	14	7.37%
public	2	2.00%	4	5.88%	6	35.29%	0	0.00%	12	6.32%
other	8	8.00%	0	0.00%	0	0.00%	0	0.00%	8	4.21%
none	27	—	—	—	—	—	—	—	27	—
totals	127	100%	68	100%	17	100%	5	100%	217	100%

EXHIBIT A-2: User Discipline Analysis: EGSUS and NASA HQ Survey (See exhibit 3-4)

USGCRP Interests of Respondents Common to EGSUS and Nasa HQ Survey (127 people)

Research Area			
Climate & Hydrologic Systems	chs	92	29.49%
Biogeochemical Dynamics	biodyn	50	16.03%
Ecological Systems and Dynamics	eco	49	15.71%
Earth System History	hist	42	13.46%
Human Interactions	hum	26	8.33%
Solid Earth Processes	solid	24	7.69%
Solar Influences	solar	29	9.29%
Total		312	100.00%

EXHIBIT A-3: User USGCRP Research Area Analysis (See exhibit 3-5)

A.2 EGSUS Products

Exhibit A-4 is a list of the 157 products available for selection by respondents to the EGSUS survey. Each product was selected at least once.

Product ID	Level	Instrument	Product Name
1A-01	1A	VIRS	VIRS Radiance
1A-PR	1A	PR	Precipitation Radar Product
1A-TMI	1A	TMI	TRMM Microwave Imager Product
1B-01	1B	VIRS	VIRS Radiance
1B-21	1B	PR	Precipitation Radar Product
1B-TMI	1B	TMI	TRMM Microwave Imager Product
1C-21	1C	PR	Precipitation Radar
2A-12	2A	TMI	TRMM Microwave Imager-Level 2A Product
2A-21	2A	PR	Precipitation Radar-Level 2A Product (Sigma Naught)
2A-23	2A	PR	Precipitation Radar-Level 2A Product (Qualitative Products)
2A-25	2A	PR	Precipitation Radar-Level 2A Product (Rainfall Profile)
2B-31	2B	TMI, PR	Level 2B Composite (TMI, and PR)
3A-11	3A	TMI	TRMM Microwave Imager-Level 3A Product
3A-25	3A	PR	Precipitation Radar-Level 3A (Monthly Structure)
3A-26	3A	PR	Precipitation Radar-Level 3A (Monthly Rainfall Map, Statistical Method)
3B-31	3B	TMI, PR	Level 3B Composite (TMI and PR)
3B-42	3B	VIRS, TMI, PR	Level 3B Composite (VIRS, TMI and PR)
3B-43	3B	VIRS, TMI, PR	Level 3B Composite (VIRS, TMI and PR)
A-01	1	RADARSAT	SCANSAR Images
A-06	1	SAR	SAR Complex Data
A-07	1	SAR	SAR Complex Data
A-08	1	SAR	SAR Images
A-09	1	SAR	SAR Images
A-10	2	SAR	SAR Geocoded Images
A-11	2	SAR	SAR Geocoded Images
A-12	3	SAR	Ice Motion Vectors
A-13	3	SAR	Ice Classification
A-14	3	SAR	Wave Spectra
A-21-25	1	RADARSAT	Complex Data Products
A-26-28	2	RADARSAT	Geocoded Images
A-32-37	1	RADARSAT	Image Data Products
A-38	2	RADARSAT	SCANSAR Terrain-Corrected Images
A-39	2	RADARSAT	Strip Mode Ice Motion Vectors
AST01	1A	ASTER	ASTER Reconstructed, Unprocessed Instrument Data
AST03	1B	ASTER	ASTER Registered Radiance at Sensor
AST04	2	ASTER	ASTER Brightness Temperature
AST05,08	2	ASTER	ASTER Land Surface Emissivity & Land Surface Temperature
AST06	2	ASTER	ASTER Decorrelation Stretch Product
AST07,09	2	ASTER	ASTER Surface Radiance and Surface Reflectance
AST13	2	ASTER	ASTER High Resolution Polar Classification
AST14	2	ASTER	ASTER Digital Elevation Models (DEMs)
CER00	0	CERES	CERES Data
CER01	1B	CERES	Bi-Directional Scan (BDS)
CER02	2	CERES	ERBE-Like Instantaneous Top Of Atmosphere and Surface Estimates (ES-8)
CER03	3	CERES	ERBE-Like Monthly Regional Averages (ES-9)
CER04	2	CERES	Single Satellite CERES Footprint, Radiative Fluxes and Clouds (CRS)
CER05	3	CERES	Hourly Gridded Single Satellite Fluxes and Clouds (FSW)
CER06	3	CERES	Monthly Top Of Atmosphere and Surface Radiation Budget Averages (SRBAVG)
CER07	3	CERES	Synoptic Radiative Fluxes and Clouds (SYN)
CER08	3	CERES	Monthly Regional Radiative Fluxes and Clouds (AVG)
CER11	2	CERES	Single Satellite Footprint, Top Of Atmosphere and Surface Flux, Clouds (SSF)
CER12	3	CERES	Hourly Gridded Single Satellite Top Of Atmosphere and Surface Fluxes/Clouds (SFC)
CER13	3	CERES	Earth Radiation Budget Experiment-Like Monthly Geographical Averages (ES-4)
CER14	3	CERES	Earth Radiation Budget Experiment-Like Monthly Gridded Averages (ES-4G)
CER15	3	CERES	Monthly Zonal and Global Radiative Fluxes and Clouds (ZAVG)
CER16	3	CERES	Clear Reflectance History (CRH)
CERX06	3	CERES	Meteorological, Ozone, and Aerosols (MOA)
COL00	0	COLOR [SeaWiFS-II]	COLOR Data
COL01	1B	COLOR [SeaWiFS-II]	COLOR Radiance Product
COL02	2	COLOR [SeaWiFS-II]	GAC Derived Product
COL03	3	COLOR [SeaWiFS-II]	Binned (Compressed) Data
COL04	3	COLOR [SeaWiFS-II]	Mapped Chlorophyll-A Product
COL05	3	COLOR [SeaWiFS-II]	Mapped K 490 Product
COL06	3	COLOR [SeaWiFS-II]	Mapped Water Leaving Radiances
COL07	3	COLOR [SeaWiFS-II]	Mapped Aerosol Radiances

COL08	3	COLOR [SeaWiFS-II]	COLOR Ancillary Data
LIS00	0	LIS	LIS Data
MIS-CCM	1B2	MISR	MISR Geometric Calibration Dataset
MIS00	0	MISR	MISR Data
MIS01	1A	MISR	MISR Product
MIS02	1B1	MISR	MISR Product
MIS03	1B2	MISR	MISR Product
MIS04	2	MISR	Top of Atmosphere and Cloud Product
MIS05	2	MISR	Aerosol/Surface Product
MIS10		MISR	MISR Ancillary Geographic Product
MIS11		MISR	MISR Ancillary Radiometric Product
MOD01	1A	MODIS	MODIS Unpacked
MOD02	1B	MODIS	MODIS Radiance, Calibrated
MOD03	1A	MODIS	Geolocational Fields
MOD04	2&3	MODIS	Aerosol Product
MOD05	2&3	MODIS	Precipitable Water
MOD06	2	MODIS	Cloud Product
MOD07	2&3	MODIS	O3 Total Burden
MOD08	2&3	MODIS	Stability (Lifted Index), Atmospheric
MOD09	2	MODIS	Surface Reflectance
MOD10	2	MODIS	Snow Cover
MOD11	2&3	MODIS	Land Surface Temperature/Emissivity
MOD12	3	MODIS	Land Cover Type
MOD13	2	MODIS	Vegetation Indices
MOD14	2	MODIS	Thermal Anomalies (Fire Size & Temperature)
MOD15	4	MODIS	Leaf Area Index & Fraction Photosynthetically Active Radiation
MOD16	3	MODIS	Evapotranspiration
MOD17	4	MODIS	Vegetation Production, Net Primary Production
MOD18	2&3	MODIS	Water-leaving Radiance
MOD19	2&3	MODIS	Pigment Concentration
MOD20	2&3	MODIS	Chlorophyll Fluorescence - Level 2 & 3
MOD21	2&3	MODIS	Chlorophyll-A Pigment Concentration
MOD22	2&3	MODIS	Photosynthetically Active Radiation
MOD23	2&3	MODIS	Suspended-Solids Concentration, Ocean Water
MOD24	2&3	MODIS	Organic Matter Concentration
MOD25	2&3	MODIS	Coccolith Concentration, Detached
MOD26	2&3	MODIS	Ocean Water Attenuation Coefficient
MOD27	2&3	MODIS	Ocean Productivity
MOD28	2&3	MODIS	Sea Surface Temperature
MOD29	2	MODIS	Sea-Ice Max Extent
MOD30	2	MODIS	Temperature and Moisture Profiles
MOD31	2&3	MODIS	Phycoerythrin Concentration - Level 2 & 3
MOD33	3	MODIS	Gridded Snow Cover
MOD34	3	MODIS	Gridded Vegetation Indices
MOD36	2&3	MODIS	Absorption Coefficient, Gelbstof & Total
MOD37	2	MODIS	Ocean Aerosol Properties
MOD40	3	MODIS	Gridded Thermal Anomalies
MOD41	2&3	MODIS	Land Surface Resistance Index
MOD42	3	MODIS	Gridded Sea-Ice Cover
MOP00	0	MOPITT	MOPITT Data
MOP01	1B	MOPITT	MOPITT Radiance
MOP02	2	MOPITT	CH4 Column (Total Burden)
MOP03	2	MOPITT	CO Profiles
MOP04	2	MOPITT	CO Column (Total Burden)
MOP05	3	MOPITT	CH4 Column (Total Burden) Gridded Product (Fourier Coefficient Form)
MOP06	3	MOPITT	CO Profiles Gridded Product (Fourier Coefficient Form)
MOP07	3	MOPITT	CO Column (Total Burden) Gridded Product (Fourier Coefficient Form)
SAG01	1B	SAGE-III	Transmission Profiles (66 wavelengths), Solar
SAG02	2	SAGE-III	Aerosol Extinction Profiles (at 7 wavelengths)
SAG03	2	SAGE-III	Cloud Height, Top
SAG04	2	SAGE-III	H2O Concentration and Mixing Ratio
SAG05	2	SAGE-III	NO2 Concentration and Mixing Ratio
SAG06	2	SAGE-III	NO3 Concentration and Mixing Ratio, Lunar
SAG07	2	SAGE-III	O3 Concentration and Mixing Ratio
SAG08	2	SAGE-III	OClO Concentration and Mixing Ratio, Lunar
SAG09	2	SAGE-III	Pressure and Temperature Profile
SWS00	0	SWS	SWS Data
SWS01	1B	SWS	SWS Backscattered Power
SWS02	2	SWS	Global Backscatter Cross Section (Sigma-Naught)
SWS03	2	SWS	Near Surface Vector Winds in Measurement Swath
V_DAO_1	4	DAO	Prognostic Sigma Level Output
V_DAO_2	4	DAO	Primary Diagnostic Sigma Level Output

V_DAO_3	4	DAO	Secondary Diagnostic Sigma Level Output
V_DAO_4	4	DAO	Surface and Vertically-Integrated Fields
V_DAO_5	4	DAO	Analysis Increments
V_DAO_6	4	DAO	Pressure Level Output
V_DAO_7	4	DAO	Analysis Fields on Pressure Levels
V_DAO_8	4	various	First Guess Minus Observations
V_ETM_L0R	0	LANDSAT 7-ETM	Level 0R Products - (radiometric and geometric corrections attached but not applied)
V_LIS_LAP	2	LIS	Lightning Area Product
V_LIS_LBP	1B	LIS	Lightning Background Images Product
V_LIS_LEP	1B	LIS	Lightning Events Product
V_LIS_LF1	4	LIS	Lightning Flash Density Product (Orbit)
V_LIS_LF2	4	LIS	Lightning Flash Density Product (15 day)
V_LIS_LF3	4	LIS	Lightning Flash Density Product (30 day)
V_LIS_LFP	2	LIS	Lightning Flash Product
V_LIS_LGP	1B	LIS	Lightning Group Product
V_LIS_LOP	2	LIS	Lightning Orbit Product
V_LIS_LVP	3	LIS	Lightning Vector Product
V_MIS_APP		MISR	Aerosol Physical and Optical Properties Product
V_MOD-0	0	MODIS	MODIS Data
V_MOD_OCD	2	MODIS	MODIS Ocean Calibration Data

EXHIBIT A-4: EGSUS Product Choices

This page intentionally left blank.

APPENDIX B: PRODUCTION MODEL ANALYSIS DETAIL

The Main Section from the CDR Technical Baseline contains the description of the various attachments (spreadsheets and text) that appear in the CDR Table of Contents (Exhibit 3-2). It is reproduced here.

Technical Baseline

This memo describes the technical baseline for the ECS engineering activities towards Release A CDR and Release B IDR. It should be noted that this baseline is not consistent with our contractual cost baseline. This memo defines a combination of files or documents that together reflect a definition of the items necessary to establish a consistent technical baseline for Release A CDR and Release B IDR. The following items are contained within the technical baseline:

- Mission Baseline (Spacecraft/Instrument manifests)
- Data Product Set (Data products/parameters and required resources - processing, storage and dependencies
- Landsat 7 and TRMM (TSDIS) requirements
- V0 Data Migration (Baseline plan for the migration of V0 data products to ECS)
- User “pull” baseline (Baseline user load in terms of number of users, accesses and distribution load for various time periods)
- Level 3 Requirements baseline (F&PRS Version and any modifications)
- M&O DAAC Implementation Baseline (DAAC activation and hours of operational support)
- Phasing of Capacities (Capacity buildup (processing/archive) relative to launch for those products not defined by the Ad hoc Working Group on Products (AHWGP) results

Changes in this baseline from the previous baseline dated Dec. 21, 1994 are:

1) The following changes were made to the Mission Baseline information:

- Changed launch date of first SAGE III instrument on the METEOR platform to Aug., 1998. The previous baseline had it’s launch date as Jan., 2000.
- Launch date of COLOR on FOO changed to July, 1998. The previous baseline launch date was May, 1998.
- The name of the Solid State Altimeter (SSALT) on RADAR ALT has been changed to the Dual Frequency Altimeter (DFA) and the Altimetry Microwave Radiometer (AMR) on RADAR ALT is changed to Microwave Radiometer (MR).

- The DORIS instrument has been removed from RADAR ALT. Previous baseline had TBD (0) capacity values.

2) Phasing of the capacities for Science Software Integration and Test and reprocessing were updated to be consistent with the AHWGP information.
3) Landsat 7 requirements are now incorporated in the F&PRS and IRD. Therefore reference to previous documention for Landsat 7 baseline has been deleted.
4) Numerous modifications were made to the files (Processing Descriptions and File Descriptions) to reflect modifications to the AHWGP data. In addition, data for other missions (TSDIS data for TRMM, Landsat 7, DAO, RADAR ALT) were added to this information. The two associated files (Processing Timelines and Volume Timelines) were updated based on these modifications. The AHWGP modifications are described in the Microsoft Word file "**changes in v2.1**".

The baseline information is being made available both on the ECS public server and through the development (limited access) partition of EDHS. This provides access by both ECS and ESDIS to this information. The baseline is described in a series of files that are available on the Public volume in a subfolder titled "Version 6/21/95" in a subfolder titled "Tech Baseline" in the "baseline" folder.

The product set baseline reflects the results of the AHWGP for the TRMM and AM-1, Landsat 7 ADEOS II (SeaWinds), and RADAR ALT missions. For the other missions, the SPSO database information as of Nov. 1, 1994 was used. This incorporates the SPSO's normalized values for processing. The spacecraft information in the SPSO data base reflects the updated mission baseline out of the program restructuring activity. Based on mutual agreement with ESDIS, one mission in the restructured baseline that is not currently reflected in this baseline is the CERES Flight of Opportunity (FOO) mission. The files included in the baseline description consist of the following:

Mission Baseline 6/21/95: Microsoft Word file providing the platforms, launch dates and instrument manifest. The CII instrument on CHEM is assumed to be a Japanese instrument with no ECS processing or archiving requirements.
Data Volume Summary 6/21/95: Summary (Excel Spreadsheet) of the data volume of the EOS products for each instrument and product level (includes both at and post-launch products. This table also shows the delta from our previous PDR baseline data volume.
Processing Summary 6/21/95: Summary (Excel Spreadsheet) of the processing load of the EOS products for each instrument and product level (includes both at and post-launch products. This table also shows the delta from our previous PDR baseline processing requirements.
The AHWGP information provided in a set of five files consisting of the following information:

AHWGP Tables—Definitions: Microsoft Word file defines the contents (columns in each table) of the two tables (Process Descriptions and File Descriptions) provided by the AHWGP.

Process Descriptions v2.1: Information (Excel Spreadsheet) provided by the AHWGP to provide the description of the processes that produce the products for CERES/LIS and all AM-1, Meteor (SAGE III), ADEOS (SeaWinds), and RADAR ALT instruments. This includes items such as number of operations, input/output files, and when the processes are executed.

File Descriptions v2.1: Information (Excel Spreadsheet) describing the input and output files for the processes defined by the AHWGP. This includes file type (e.g. temporary, archive), sizes, archive site and a reference to the associated SPSO products where applicable. File sizing for Level 0 data is also not included.

Processing Timelines v2.1: File (Excel Spreadsheet) that reflects the processing load (Mflops) that are required for each instrument/process that are derived from the AHWGP information. The information is provided for each calendar quarter.

Volume Timelines v2.1: File (Excel Spreadsheet) that reflects the volume (Gbytes/day) that are required to be archived for each file that are derived from the AHWGP information. The information is provided for each calendar quarter. This is only for files that must be archived (e.g., temporary/interim files are not included).

Product Details v1.2U (Norm): Detail information (Excel Spreadsheet) on each EOS product from the SPSO database. The information consists of the following for each product: number of parameters, production and archiving DAAC, indicator whether at or post launch, level, indicator if standard or special product, indicator whether routine or on-request, and the data volume (GB/day) and processing requirements (MFLOPS). The new baseline is the normalized values for Nov 1, 1994 - Columns R and T in the table for volume and processing respectively. The information in this file should be used for all the products of all instruments not included in the AHWGP information. This would be for all instruments beyond AM-1 except for CERES on PM-1. In addition it can be used as a cross reference for additional information to the AHWGP data but where there is any conflict the AHWGP information has precedence.

Parameter Details : Detail information (Excel Spreadsheet) for each EOS parameter from the SPSO database. The information includes the following for each parameter: Product ID, Parameter name and ID, investigator, Units, accuracy, temporal and spatial resolution, parameter volume and processing requirements, input data requirements (product dependencies) for product generation and communications required for QA. Similar to the Product Details, The information in this file should be used for all the products/parameters of all instruments not included in the AHWGP information. This would be for all instruments beyond AM-1 except for CERES on PM-1. In addition it can be used as a cross reference for additional information to the AHWGP data but where there is any conflict the AHWGP information has precedence.

Other information that describes the current technical baseline consists of the following:
TRMM: included in the "Tech Baseline" folder on public is a Microsoft Word file titled "**TSDIS Products**" that provides the volumes of the TSDIS products for each product level. More detailed information for the products that make up each of the product levels is available from the library in document LIB2088 titled "**T**ropical Rainfall Measuring Mission Science Data and Information System Product Volume Estimates" dated June 27, 1994.
V0 Data Migration: The baseline V0 migration plan is included in the "Tech Baseline" folder on public in a Microsoft Word file titled "**V0 Migration Baseline 6/21/95**". This information describes the phasing of the migration of the V0 data sets to ECS.
User "pull" Baseline: The user "pull" baseline is available in the "Tech Baseline" folder on public in a Microsoft Word file titled "**User Pull Baseline 6/21/95**". This describes the anticipate user load on ECS for various time periods. The system load information includes the anticipated number of users, system accesses and volume to data to be distributed. An additional Excel spreadsheet file titled "**DAAC Pull Baseline 6/21/95**" has been included to provide the allocation of the user pull load to each of the DAACs. Addition information reflecting the answers to specific user pull questions from the system designers is available in the Microsoft Word file titled "**User Pull Info 6/21/95**".
L3 Requirements Baseline: The L3 requirements baseline is available in the "Tech Baseline" folder on public in a Microsoft Word file titled "**Reqts Baseline 6/21/95**". This describes the baseline version Functional and Performance Requirements Specification (F&PRS) and any modifications to it we are assuming for the current set of Level 3 requirements.
M&O DAAC Implementation Baseline : the baseline M&O DAAC implementation baseline is available in the "Tech Baseline" in two MacSchedule files titled "**DAAC Imp Baseline pt1**" and "**DAAC Imp Baseline pt2**". This describes the time period for the activation of each of the DAACs along with the installation periods for the ECS releases. Also provided is the hours of operations of each of the DAACs versus time. This is also available in hardcopy form from the library in document LIB2110.
Phasing of Capacities: the baseline for the phasing of processing and archive capacities to support a ramp-up to standard products and archiving. This information reflects a phasing of the capacities relative to the launch date of each spacecraft. This information is available in a Microsoft Word file in the "Tech Baseline" folder titled "**Capacity Phasing 6/21/95**". The phasing specified in this file is used to phase in the capacitites for Science Software I&T and reprocessing. This would include TSDIS, Landsat 7 and EOS missions beyond AM-1. The phasing of reprocessing to support the products covered by AHWGP is being developed by Bowyer, Endal and Daly.

APPENDIX C: Performance Model Analysis Detail

This appendix contains the details of the performance model evaluation. The categories in the exhibits correspond to those in Section 5 but are expanded into a lower level of detail. The exhibits containing the evaluation results are organized into the four major evaluation areas described in Section 5. Exhibit C-1 shows how the evaluation areas are decomposed into lower levels for evaluation purposes. The following notes apply to all exhibits in the performance model evaluation.

Notes:

Analysis Priority:
1=highest, 3=lowest

Level of Maturity:
0=not addressed
1=limited maturity
2=somewhat limited maturity
3=fully mature
Blank=not yet analyzed by IV&V

Analysis Date:
Date that listed level of maturity was achieved, as reported in a TAR or TAM.

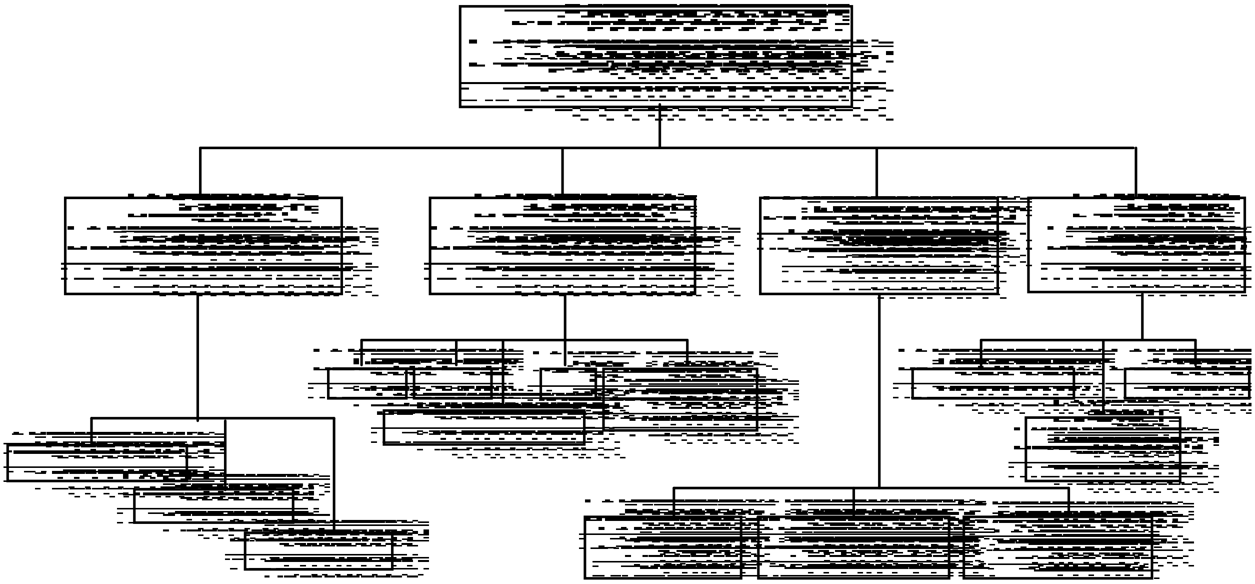


EXHIBIT C-1: Evaluation Decomposition

C.1 System Representation Evaluation

The System Representation evluation is contained in exhibits C-2 through C-4 below.

Subsystem	Analysis Priority	Level of Maturity	Analysis Date
Client	3		06/30/95
Interoperability	1	2	06/30/95
Data Management	1	0	06/30/95
Data Server	1	3	06/30/95
Ingest	1	3	06/30/95
Planning	1	2	06/30/95
Data Processing	1	3	06/30/95

EXHIBIT C-2: Subsystem Evaluation

Resource Category	Analysis Priority	Level of Maturity	Analysis Date
Processors	1	2	06/30/95
Disks	1	2	06/30/95
R/W Heads	1	3	06/30/95
Robots	1	3	06/30/95
I/O Channels	2	2	06/30/95
Network Links	1	2	06/30/95
Memory	2	0	06/30/95

EXHIBIT C-3: Resource Evaluation

Subsystem Interfaces	Analysis Priority	Level of Maturity	Analysis Date
Ingest	1	3	06/30/95
Interoperability	1	2	06/30/95
Data Server	1	3	06/30/95
Client	3		06/30/95
Planning	1	2	06/30/95
Data Management	1	0	06/30/95
Data Processing	1	3	06/30/95

EXHIBIT C-4: Subsystem Interface Evaluation

C.2 Workload Representation Evaluation

The Workload Representation evluation is contained in exhibits C-5 through C-9 below.

Instrument	Analysis Priority	Level of Maturity	Analysis Date
ACRIM	3		06/30/95
AIRS	3		06/30/95
AMSU	3		06/30/95
ASTER	1	3	06/30/95
CERES (TRMM)	1	3	06/30/95
CERES (AM1)	1	3	06/30/95
DORIS	3		06/30/95
ESOP	3		06/30/95
ETM	3		06/30/95
GLAS	3		06/30/95
GV	1	3	06/30/95
HIRDLS	3		06/30/95
LIS	1	3	06/30/95
MHS	3		06/30/95
MIMR	3		06/30/95
MISR	1	3	06/30/95
MLS	3		06/30/95
MODIS	1	3	06/30/95
MOPITT	1	3	06/30/95
PR	1	3	06/30/95
SAGE-III	3		06/30/95
SeaWiFS	3		06/30/95
SOLTICE	3		06/30/95
SSALT	3		06/30/95
SWS	3		06/30/95
TES	3		06/30/95
TMI	1	3	06/30/95
TMR	3		06/30/95
VIRS	1	3	06/30/95

EXHIBIT C-5: Push Workload Evaluation

Pull Workload	Analysis Priority	Level of Maturity	Analysis Date
Simple Search, 1 Site	1	0	06/30/95
Simple Search, M Sites	1	0	06/30/95
Match-Up Search, 1 Site	1	0	06/30/95
Match-Up Search, M Sites	1	0	06/30/95
Coincident Search, 1 Site	1	0	06/30/95
Coincident Search, M Sites	1	0	06/30/95
Archive (Insert)	1	0	06/30/95
Ingest	1	2	06/30/95
Inspect (Browse)	1	0	06/30/95
Produce	1	0	06/30/95
Exchange	1	2	06/30/95
Standing Order (Subscription)	1	0	06/30/95
Manipulate (Subset)	1	2	06/30/95
Modify	1	0	06/30/95
Acquire (Exchange)	1	0	06/30/95

EXHIBIT C-6: Pull Workload Evaluation

Instrument	Analysis Priority	Level of Maturity	Analysis Date
ACRIM	3		06/30/95
AIRS	3		06/30/95
AMSU	3		06/30/95
ASTER	1	3	06/30/95
CERES (TRMM)	1	0	06/30/95
CERES (AM1)	1	3	06/30/95
DORIS	3		06/30/95
ESOP	3		06/30/95
ETM	3		06/30/95
GLAS	3		06/30/95
GV	1	3	06/30/95
HIRDLS	3		06/30/95
LIS	1	3	06/30/95
MHS	3		06/30/95
MIMR	3		06/30/95
MISR	1	3	06/30/95
MLS	3		06/30/95
MODIS	1	3	06/30/95
MOPITT	1	3	06/30/95
PR	1	3	06/30/95
SAGE-III	3		06/30/95
SeaWiFS	3		06/30/95
SOLTICE	3		06/30/95
SSALT	3		06/30/95
SWS	3		06/30/95
TES	3		06/30/95
TMI	1	3	06/30/95
TMR	3		06/30/95
VIRS	1	3	06/30/95

EXHIBIT C-7: Reprocessing Workload Evaluation

Category	Analysis Priority	Level of Maturity	Analysis Date
Data Migration	2	3	06/30/95
V0 User Access	2	0	06/30/95

EXHIBIT C-8: V0 Workload Evaluation

System Overheads	Analysis Priority	Level of Maturity	Analysis Date
Workload Overhead	2	1	6/30/95
Resource Overhead	2	1	6/30/95

EXHIBIT C-9: System Overheads

C.3 Performance Statistics Evaluation

The Performance Statistics evluation is contained in exhibits C-10 through C-12 below.

Resource	Analysis Priority	Level of Maturity	Analysis Date
Processor	1	3	6/30/95
Disk	1	2	6/30/95
I/O Channel	3	0	6/30/95
Robot	1	3	6/30/95
R/W Head	1	3	6/30/95
Network	1	3	6/30/95
Memory	1	0	6/30/95

EXHIBIT C-10: Resource Metrics Evaluation

Requirement Category	Analysis Priority	Level of Maturity	Analysis Date
System	1	0	06/30/95
PGS	1	0	06/30/95
DADS	1	0	06/30/95
IMS	1	0	06/30/95
ESN	1	0	06/30/95

EXHIBIT C-11: Performance Requirements Evaluation

Results Category	Analysis Priority	Level of Maturity	Analysis Date
Results Validation	1	2	06/30/95
Analysis Methods	1	3	06/30/95

Trade Studies	1	2	06/30/95

EXHIBIT C-12: Performance Results Evaluation

C.4 Model Structure Evaluation

The Model Structure evluation is contained in exhibits C-13 through C-15 below.

Module Category	Analysis Priority	Level of Maturity	Analysis Date
Workloads	2	3	06/30/95
Statistics Collection	2	3	06/30/95
Subsystems	2	3	06/30/95

EXHIBIT C-13: Module Modules Evaluation

Data Structure Category	Analysis Priority	Level of Maturity	Analysis Date
Workloads	2	3	06/30/95
Processes	2	3	06/30/95
Data	2	3	06/30/95
Configuration	2	3	06/30/95

EXHIBIT C-14: Model Data Structures Evaluation

Parameter Categories	Analysis Priority	Level of Maturity	Analysis Date
Workload Frequencies	2	3	06/30/95
Resource Configuration	2	3	06/30/95
Resource Capacities	2	3	06/30/95
Process Characteristics	2	3	06/30/95
Data Characteristics	2	3	06/30/95

EXHIBIT C-15: Model Parameters Evaluation

This page intentionally left blank.

APPENDIX D: COST MODEL ANALYSIS DETAIL

D.1 Analysis Methods

It was the intent of the IV&V team to perform a complete cost analysis for this report. When it became obvious that we would not be provided access to the models and most of the supporting cost information, we choose to follow an investigative path. We started by collecting as much relevant data as we could from the our interviews and correspondence with the model developers. We then compared that information against similar cost information and other modeling efforts. This appendix contains additional details related to our cost modeling analysis.

D.1.1 COTS HW and SW Estimation

In the area of COTS hardware and software, there are a number of specifics that tend to drive costs. This includes the numbers of components, the cost per component, and cost trends. Regarding the number of components, the key issue addressed was how the numbers were derived. This in itself is a complex issue. There was no attempt in this analysis to duplicate the work reported in Section 5. Rather, the goal was to determine whether performance modeling was used, and if not, what techniques were used. For component cost, the obvious key issue is whether the costs used for specific components are similar to those available in the marketplace. Given the historical increase in performance and reduction in cost as a function of time, and the fact that any major system, such as EOSDIS is always built over a period of time, the key issue with cost trends is whether that historical performance improvement/cost reduction has been accounted for in estimating costs.

Regarding the COTS cost estimation model, the key issue was whether the model supports its primary purpose; i.e., trade-off analysis. Speed and ease of use are key to achieving that purpose. In contrast, regarding the COTS procurement model, the key issue was its ability to produce accurate costs. For this model, speed and ease of use are less of an issue. Attention must be paid to details regarding the specific components used, their numbers, and their costs.

A variety of methods were used to perform the Cost Model evaluation. The initial plan was to become familiar with the models, their purposes, and the general way they were implemented, and then obtain the models and examine them directly. When it became clear that the models were not going to be made available, the analysis became more focused on learning about the models and performing the evaluation on the basis of that knowledge. Hence, the evaluation was performed by:

- Reviewing existing briefing materials
- Interviewing the model developer
  - conduct interviews
  - produce findings
  - ask follow-up questions

COTS Cost Estimation Model - This model is a cost estimation tool that provides an estimated cost for COTS hardware and software, associated maintenance, and operations costs associated with the COTS hardware through the end of the contract. The model provides a decomposition of the system costs into element costs and the ability to estimate costs for processing hardware and software.

This model decomposes the system based upon the design submitted as part of the original HAIS proposal. It generally follows the decomposition represented in the F&PRs, i.e. SDPS (PGS, DADS, IMS) , CSMS (SMC/LSM, ESN), etc. However, it decomposes the system a step further into sub-elements. For DADS, this includes archive, ingest, and distribution.

This model provides the capability to estimate costs for processing hardware based upon the MFLOPS estimates (such as those from the SPSO database), and RAID disk and storage archive based upon the data volume estimates (also available from the SPSO database). The cost of each sub-element can be estimated by one of at least four ways. The method used for each sub-element is controlled by the operator. The four methods are described below:

a) Based on data volumes / MFLOPS from the SPSO database or the AHWGP.
b) Based on the following formula: technical baseline data volume / SOW data volume * cost from proposal for the element and sub-element in question.
c) Based on the following formula: technical baseline granule volume / SOW granule volume * cost from proposal for the element and sub-element in question.
d) Added directly from contract or Change Order #1. (Note that Change Order #1 cost additions were derived by costing the added configuration items using the COTS Procurement Cost Model).

In the case of methods b) and c), the technology which formed the basis for the contract cost is used as the basis of the estimate.

COTS Procurement Model - The COTS procurement cost model, also known as the Bill-of-Materials Procurement Cost Model, is a means of estimating COTS hardware and software costs given a bill-of-materials. Inputs are provided to this model in the form of code names for each item in the bill-of-materials, the required quantity, and the release for which the items must be procured. Given this set of inputs, the model equates the code name to a specific make and model of hardware or software, converts the release entry to a date, applies phasing, then applies a cost as a function of time curve. No information regarding the rates of decrease of price per unit performance used within this model was obtained. The output of the model is the total COTS hardware and software cost across the life of the project for the input bill-of materials. This model also has the capability to compute maintenance costs as a function of time.

D.1.2 Custom SW Estimation

The custom software estimation analysis focused on studying the method HAIS used to arrive at their software size estimation. This estimating methodology utilized the output from the size estimate as the primary input for the schedule and cost estimate.

A complete analysis to verify the HAIS methodology could not be performed due to a lack of information. The majority of the effort was focused on assessing the engineering quality of the estimation process. Software estimation follows four basic steps:

- size estimation
- level of effort estimation
- schedule determination
- cost estimation

Due to lack of available information, the analysis focused on determining how the steps mentioned above were performed. It was determined early in the evaluation process that a somewhat new method was being used to estimate software size. Therefore, a key aspect of the analysis addressed the validity of that method. Some of the specific tasks performed included:

- Determining what methods exist for estimation of software size/effort given an object-oriented design process and the degree to which they have been tested
- Searching for an established relationship between source lines of code (SLOC) and object-oriented design (OOD) entities;
- Determining the relationship between SLOC and OOD entities such as objects and methods in existing C++ software

As discussed in Section 6.1 and 6.3, even if a valid method is used, the parameters used within the method also strongly effect the accuracy of the estimated costs. Hence, this analysis examined three key parameters: numbers of elements, size multipliers, and cost trends. The numbers of elements in this context refer to counts of object-oriented design entities, such as objects (similar to structures in C) and methods (similar to functions in C). Size multipliers refer to the number of SLOC used to multiply against the counts of design entities. Cost trends in this context refer to whether changes in cost (personnel cost in this case) were recognized and accounted for.

The other issues included in evaluation of technical integrity were similar to those discussed in Section 6.1 and 6.3.

Throughout the analysis no hard copy materials were provided to aid in the evaluation. Since the estimation approach being used is somewhat new, the analysis did include reviewing the literature available on techniques used for estimating OOD software.

Our research did reveal that there are a number of software estimating programs available which can assist in translating a LOC estimate into an object oriented design. Further, there are full-scale lifecycle estimating packages available capable of creating an estimated cost for software creation, as well as a detailed development plan which estimates the effort and produces an integrated schedule. Those programs are:

- PRICE® S, by Martin Marietta Price Systems
700 East Gate Drive, Suite 200
Mount Laurel, New Jersey 08054
(609) 866-6789

- SEER™-SEM, by Galorath Associates, Inc.
9920 S. La Cienega Blvd.
LAX (Inglewood), CA 90301
(310) 670-3404

- SOFTSTAR, by Softstar Systems
28 Ponemah Road
Amherst, NH 03031
(603) 672-0987

D.1.3 Operational Cost Estimation

Operational costs are driven by personnel costs. Therefore, the key issues regarding estimation of operational costs center around how personnel costs are estimated. Creating a complete estimate requires that all necessary staff are accounted for in the estimate. Conversely, the estimate must insure that the personnel included are required.

The engineering quality of the operation estimate related to method used to estimate the numbers of personnel and if the schedule for delivery of automated features was factored into the estimation process.

Early in the analysis it was determined that a fixed value of $100K was being assumed for the cost of a man-year. Because no hard data was available, in order to evaluate the validity of using such a figure, a typical mix of labor categories and costs per labor category was assumed. The average salary resulting from these assumptions was then computed and compared to the stated value of $100K. The inputs to this analysis of average annual labor cost are shown in exhibit D-2.

POSITION TITLE	**Number of Staff**	**Cost/HR (unloaded)**	**Multiplier used**	**ANNUAL COST**	**TOTAL COST**
M&O Manager	4	$40	off-site	$188,416	$753,664
Site Manager	1	$35	on-site	$143,360	$143,360
Admin Support / Security	3.7	$15	on-site	$61,440	$227,328
Librarian	1	$15	on-site	$61,440	$61,440
Operational Readiness and Performance Assurance	2	$20	on-site	$81,920	$163,840
DAAC Trainers	2	$20	on-site	$81,920	$163,840
SUSTAINING ENGINEERING					
S/W Maintenance/Engineering	20.2	$30	off-site	$141,312	$2,854,502
Science	2	$25	on-site	$102,400	$204,800
Planned Upgrades	1.7	$25	on-site	$102,400	$174,080
Configuration Management	7.2	$25	on-site	$102,400	$737,280
Testing	3	$25	on-site	$102,400	$307,200
Property Management/ILS	4.4	$15	on-site	$61,440	$270,336
H/W Maintenance	10	$30	on-site	$122,880	$1,228,800
Resource Controller/Performance Analyst	1	$25	on-site	$102,400	$102,400
ALGORITHM SUPPORT					
Test & Integration	6	$20	on-site	$81,920	$491,520
Development	4	$25	off-site	$117,760	$471,040
Data Base Administration	2	$20	on-site	$81,920	$163,840
Ops. Supervisor/Production Scheduler	1	$20	on-site	$81,920	$81,920
QA/Production Monitor	22.1	$20	on-site	$81,920	$1,810,432
Ground Controller	7.2	$20	on-site	$81,920	$589,824
USER SERVICES					
Data Specialist	10.2	$20	on-site	$81,920	$835,584
User Assistance	13	$15	on-site	$61,440	$798,720
Data Distribution Technician	8.2	$15	on-site	$61,440	$503,808
Computer Operator	10.3	$15	on-site	$61,440	$632,832
Archive Manager	4.7	$20	on-site	$81,920	$385,024
TOTALS	151.9				$14,157,414
		Average cost per man-year =			**$93,202**
onsite_indirect_rate	2				
offsite_indirect_rate	2.3				

EXHIBIT D-1: Parameters Used in Estimating the Average Cost per Man Year

D.2 Analysis Results

The results of analysis of traceability to requirements is shown in exhibit D-2.

REQUIREMENT	**SATISFIED?**	**COMMENTS**
The contractor shall establish and maintain a Life Cycle Cost (LCC) model	PARTIAL	There is no overall model which computes lifecycle cost. Rather, outputs from a series of models must be manually accumulated.
The LCC model shall be developed to be compatible with the ECS Work Breakdown structure (WBS).	TBD	
The LCC model shall identify life cycle costs including the cost of development, acquisition, operation, COTS licensing, upgrades (including newer versions of COTS software), correction of latent defects, and related system support over the ECS lifetime.	PARTIAL	The components of cost described in the requirement are being estimated individually. There is, however, no overall model which rolls up these components of cost into an overall lifecycle cost.
The LCC model shall also include the cost of any necessary maintenance subcontracts.	YES	
The LCC model shall include projections for technology improvements.	YES	
The contractor shall provide ECS Life-Cycle Cost Reports in accordance with DID 213/SE2.	TBD	
The LCC model shall model cost sensitive parameters to provide the Government with the capability assess cost and schedule impacts of new or modified requirements. Cost sensitive parameters shall include, but not be limited to: new instruments, schedule changes, processing requirements, archive volume requirements, number of granules, number of products, and input/output loads.	PARTIAL	View #1: Since there is no standalone model, there is no capability for the government to do this. View #2: The "Interactive Cost Model" partially fulfills this requirement.
The LCC Model shall be continuously updated with actual performance data.	YES	Changes have been made between PDR and CDR.
The LCC Model, as well as the results from it, shall be made available to the Government.	PARTIAL	Some parts of model have been delivered. However, the entire model, cannot be made available, since it does not exist in standalone form.

EXHIBIT D-2: Cost Modeling Requirements Satisfaction Matrix

This page intentionally left blank.

APPENDIX E: LIST OF REFERENCES

The following documents were referenced in the assessment of the ECS models or in the preparation of this report.

1. *EOSDIS Modeling Assessement Plan*, EOSDIS IV&V, Intermetrics, Greenbelt, MD, September 14, 1994
2. *EOSDIS Modeling Assessment Report (Draft Preliminary)*, EOSDIS IV&V Deliverable #IVV-0506, Intermetrics, Greenbelt, MD, October 7, 1994
3. *EOSDIS Core System (ECS) Modeling Assessment Report (Preliminary),* EOSDIS IV&V, Deliverable 0506, Intermetrics, Greenbelt, MD, February 10, 1995
4. *Technical Baseline for the ECS Project*, Technical Paper 210-TP-001-003, Hughes Information Technology Company, Landover, MD, June 21, 1995
5. *User Characterization and Requirements Analysis,* White Paper 194-00312TPW, Hughes Information Technology Company, Landover, MD, September 1994
6. *Systems Performance Model for the ECS Project*, Technical Paper 241-TP-001-001, Hughes Information Technology Company, Landover, MD, February 1995
7. *Science Data Processing System (SDPS) Segment Design Specifications for the ECS Project*, 305-CD-002-002, Hughes Information Technology Company, Landover, MD, March 1995
8. *Interactive Cost Model*, Technical Analysis Memorandum EOSVV-TAM-05-001-3/27/95, EOSDIS IV&V, Intermetrics, Greenbelt, MD, March 27, 1994
9. *Interactive Cost Model - A Scientist's View,* Technical Analysis Memorandum EOSVV-TAM-05-002-3/27/95, EOSDIS IV&V, Intermetrics, Greenbelt, MD, March 27, 1994

This page intentionally left blank.

APPENDIX F: TOOLS AND DATA BASES UTILIZED

This section describes the data bases and tools used by IV&V team to assess the User, Production, Performance and Cost Models. Specific data bases and tools used, including specific version and operational environment, are provided in Exhibit F-1.

IV&V TOOLS	ENVIRONMENT	SPECIFIC DATABASES	MODEL
MS Excel v5.0	PCs	HAIS EOSDIS User Scenarios CDR Technical Baseline SPSO Science Data Plan SPSO Output Data Products & Input Req.	User Production Production Production
MS Word v6.0c	PCs		All
MS Access	PCs		All
MS PowerPoint	PCc		All
Lotus Approach v3.0	PCs	MTPE Data Base EGSUS Database	User User
BONeS 2.6.1	SUN, VT 110, 220 (Emulator)		Performance Production
Novell Netware LAN/WAN LAN WorkPlace Internet	PCs SUN, PCs, and Mac SUN, PCs SUN, PCs, Mac		All

EXHIBIT F-1: Tools and Databases Used

In addition, the IV&V Cost Model analyses utilized several specialized external tools and databases:

COTS H/W and S/W Estimation: External sources were consulted to determine the realistic trends in price versus performance. For CPU performance vs. time, *Business Week*, July 4, 1994 and NETLIB at Oak Ridge National Labs were consulted. For archive capacity vs. time, the National Media Labs Independent Report was utilized.

This page intentionally left blank